Three Essays on the Econometric Analysis of High-Frequency Data

DISSERTATION

zur Erlangung des akademischen Grades
doctor rerum politicarum
(Doktor der Wirtschaftswissenschaft)

eingereicht an der
Wirtschaftswissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von
Dipl.-Vw. Peter Malec

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Jan-Hendrik Olbertz

Dekan der Wirtschaftswissenschaftlichen Fakultät:
Prof. Dr. Ulrich Kamecke

Gutachter:

1. Prof. Dr. Nikolaus Hautsch
2. Prof. Dr. Melanie Schienle
3. Prof. Asger Lunde, Ph.D.

Eingereicht am: 12. Februar 2013
Tag des Kolloquiums: 30. Mai 2013

Abstract

In three essays, this thesis deals with the econometric analysis of financial market data sampled at intraday frequencies. In particular, we focus on the dynamic and distributional modeling of high-frequency or intraday data, as well as its utilization for risk reduction in vast-dimensional portfolios.

Chapter 1 presents a novel approach to model serially dependent positive-valued variables realizing a nontrivial proportion of zero outcomes. This is a typical phenomenon in financial high-frequency time series. We introduce a flexible point-mass mixture distribution, a tailor-made semiparametric specification test and a new type of multiplicative error model (MEM). Applying the proposed methodology to high-frequency cumulated trading volumes of liquid and illiquid NYSE stocks, we show that the model captures the dynamic and distributional properties of the data and is able to correctly predict future distributions.

Chapter 2 addresses the problem that fixed symmetric kernel density estimators exhibit low precision for positive-valued variables with a large probability mass near zero, which is common in high-frequency data. We show that gamma kernel estimators are superior, while their relative performance depends on the specific density and kernel shape. We suggest a refined gamma kernel and a data-driven method for choosing the appropriate type of gamma kernel estimator. In a simulation study, we compare the refined estimator to the original gamma kernels and standard boundary-correction methods, demonstrating the superiority of the new approach.

Chapter 3 turns to the open debate about the merits of high-frequency data in large-scale portfolio allocation. We consider the problem of constructing global minimum variance portfolios based on the constituents of the S&P 500. Covariance matrix predictions are obtained by applying a blocked realized kernel estimator, along with different smoothing windows, regularization methods and forecasting models. We show that forecasts based on high-frequency data can yield a significantly lower portfolio volatility than approaches using daily returns, implying noticeable utility gains for a risk-averse investor.

Zusammenfassung

Diese Dissertation behandelt die ökonometrische Analyse von hochfrequenten Finanzmarktdaten. Insbesondere liegt der Schwerpunkt auf der Modellierung der Dynamik und Verteilung von Hochfrequenz- oder Intratagesdaten sowie der Ausnutzung dieser Daten für die Risikominderung in hochdimensionalen Portfolios.

Kapitel 1 stellt einen neuen Ansatz zur Modellierung von seriell abhängigen positiven Variablen, die einen nichttrivialen Anteil an Nullwerten aufweisen, vor. Letzteres ist ein weitverbreitetes Phänomen in hochfrequenten Finanzmarktzeitreihen. Eingeführt wird eine flexible Punktmassenmischverteilung, ein maßgeschneiderter semiparametrischer Spezifikationstest sowie eine neue Art von multiplikativem Fehlermodell (MEM). Die Anwendung der vorgeschlagenen Methoden auf hochfrequente kumulierte Handelsvolumina von liquiden und illiquiden NYSE-Aktien zeigt auf, dass das Modell die Eigenschaften der Daten bzgl. Dynamik und Verteilung erfasst sowie eine korrekte Prognose zukünftiger Verteilungen ermöglicht.

Kapitel 2 beschäftigt sich mit dem Umstand, dass feste symmetrische Kerndichteschätzer eine geringe Präzision aufweisen, falls eine positive Zufallsvariable mit erheblicher Wahrscheinlichkeitsmasse nahe Null gegeben ist. Wir legen dar, dass Gammakernschätzer überlegen sind, wobei ihre relative Präzision von der genauen Form der Dichte sowie des Kerns abhängt. Wir führen einen verbesserten Gammakernschätzer sowie eine datengetriebene Methodik für die Wahl des geeigneten Typs von Gammakern ein. In einer Simulationsstudie vergleichen wir den verbesserten Schätzer mit dem ursprünglichen Gammakern sowie mit Korrekturmethoden für feste symmetrische Kerne. Dabei zeigen wir die Überlegenheit des neuen Ansatzes auf.

Kapitel 3 wendet sich der offenen Frage nach dem Nutzen von Hochfrequenzdaten für hochdimensionale Portfolioallokationsanwendungen zu. Wir betrachten das Problem der Konstruktion von globalen Minimum-Varianz-Portfolios auf der Grundlage der Konstituenten des S&P 500. Die Kovarianzmatrixprognosen beruhen auf der Anwendung des geblockten Realized-Kernel-Schätzers sowie verschiedenen Glättungsfenstern, Regularisierungsmethoden und Prognosemodellen. Wir zeigen auf, dass Prognosen, welche auf Hochfrequenzdaten basieren, im Vergleich zu Methoden, die tägliche Renditen verwenden, eine signifikant geringere Portfoliovolatilität implizieren. Letzteres geht mit spürbaren Nutzengewinnen aus der Sicht eines Investors mit hoher Risikoaversion einher.

Acknowledgment

I owe a great deal of debt to many people whose help enabled me to write and, most importantly, finish this thesis.

First and foremost, I am very grateful to my supervisors, Nikolaus Hautsch and Melanie Schienle, for the extraordinary mentoring and support they have given me during the past years. The numerous discussions we had were not only exceptionally interesting and productive, but also helped me to develop and deepen my research interests. I would also like to express my gratitude to Asger Lunde for kindly accepting the task of becoming co-examiner of this thesis.

My scientific work has profited vastly from the enjoyable and fruitful research atmosphere at the chair of econometrics at Humboldt-Universität zu Berlin. For constructive comments, valuable suggestions and stimulating discussions on various subjects, I would like to thank my colleagues Gökhan Cebiroglu, Bernd Droge, Axel Groß-Klußmann, Jonas Haase, Gustav Haitz, Ruihong Huang, Lada Kyj, Franziska Lottmann, Tomas Polak and Julia Schaumburg.

When writing the present thesis, I was research assistant at the interdisciplinary Collaborative Research Center 649 "Economic Risk" supported by the Deutsche Forschungsgemeinschaft. This environment allowed for an intellectual exchange with researchers from other fields, which was both inspiring and had a considerable impact on my work. In particular, I am grateful to the team members of the Research Data Center who provided me with the necessary computing power, as well as access to all relevant databases.

Finally, I owe more than can be said to the support and encouragement offered to me by my family and friends.

Contents

List of Figures

List of Tables

Introduction

Beginning in the 1990s, one of the major developments in econometrics has been related to the analysis of financial data sampled at ever higher frequencies. This evolution was triggered by an increasing intraday trading activity, as well as advances in the technology for recording, storing and processing vast datasets. On the one hand, so-called *intraday* or *high-frequency data* offer great opportunities. The latter are due to the substantial advantage in terms of the amount of information provided when compared to, e.g., daily observations of financial variables. As a matter of fact, the limiting case is a situation in which records on every single transaction or even every order event occurring in a limit order book are available.[1] On the other hand, researchers face numerous challenges, since high-frequency data exhibit features that are not encountered at lower frequencies. Hence, its analysis requires new tailor-made econometric methods.

This thesis covers two general topics in the field of *high-frequency econometrics*. Chapter 1 and 2 are related to the problem of both *modeling* high-frequency data and *evaluating* the resulting specifications. Econometric models for variables such as high-frequency trading volumes or volatilities are crucial, e.g., for intraday trading strategies. In this context, dynamic specifications have to account for the main stylized facts of the aforementioned type of data, including a strong persistence, intraday seasonality effects and the nonnegativity of many trading variables. Chapter 1 presents an extension of the most widely-used modeling framework addressing an additional problem that is often encountered in high-frequency data sampled at a regular grid: the occurrence of a considerable number of zero observations in the dataset. The latter effect can be captured by an appropriate adjustment of the model for the underlying distribution, while a powerful tool for examining the validity of the distributional assumptions is given in the form of kernel-based specification tests. This evaluation methodology involves non- or semiparametric density estimation, which is a nontrivial task if the variable of interest is positive-valued and its distribution features a considerable probability mass close to zero. Motivated by the fact that the above properties are of empirical relevance in high-frequency datasets, Chapter 2 compares alternative estimators

[1] For the former case, Engle (2000) has introduced the term "ultra-high-frequency data".

and proposes refinements of existing approaches.

As the second major topic of this thesis, Chapter 3 discusses the *utilization* of high-frequency data for economic applications in a low-frequency setting with the focus being on portfolio selection. The underlying rationale is that covariance matrix estimates and forecasts based on high-frequency returns should offer an increased precision compared to their counterparts employing daily data. Then, the question arises whether the aforementioned precision gains also translate into economic benefits, such as a reduced portfolio risk. So far, this issue has only been addressed for a relatively moderate number of assets, while the realistic, vast-dimensional scenario has remained unexamined. Here, the main econometric challenges include an efficient estimation of the large-scale covariance matrices, ensuring that the resulting forecasts can be inverted in a numerically stable way, as well as imposing suitable smoothing schemes to prevent excessive transaction costs.

In the following, the most important contributions and findings of the three chapters that comprise this thesis will be outlined in more detail.

Chapter 1 is joint work with my supervisors, Nikolaus Hautsch and Melanie Schienle, and is forthcoming in the *Journal of Financial Econometrics*. In this chapter, we propose a new econometric methodology for modeling positive-valued variables which exhibit a strong serial dependence and, in particular, realize a non-trivial proportion of zero outcomes. The latter is a typical phenomenon in financial time series sampled at high frequencies and aggregated over regular intervals, such as cumulated trading volumes. Our proposed approach comprises both a flexible point-mass mixture distribution and a semiparametric specification test that is explicitly tailored for distributions of this type. Further, we introduce a novel type of multiplicative error model (MEM) relying on the above zero-augmented distribution. This structure includes an autoregressive binary choice component and thus is able to capture the (potentially different) dynamics of zero occurrences and of strictly positive realizations. In an empirical study, we apply the proposed modeling framework to high-frequency cumulated trading volumes of both liquid and illiquid NYSE stocks. Our results demonstrate that the model accounts for the dynamic and distributional features of the data and provides precise forecasts of future distributions.

Chapter 2 is joint work with my supervisor Melanie Schienle. A version of this chapter is being revised for publication in *Computational Statistics & Data Analysis*. The chapter tackles the known issue that standard fixed symmetric kernel-type density estimators suffer from precision problems when considering positive-valued random variables with a pronounced probability mass close to the lower bound of the support. In particular, we demonstrate that, in such settings, alternative approaches based on asymmetric gamma kernel estimators are superior. However,

both the asymptotic and finite sample performance of the aforementioned estimators differ depending on the shape of the underlying density near zero, as well as the specific form of the chosen kernel. Hence, we propose a refined gamma kernel featuring an additional tuning parameter. The latter is varied according to the shape of the density near the lower bound of the support. In addition, we suggest a data-driven methodology which allows to choose the modified gamma kernel estimator appropriately. We conduct a comprehensive simulation study comparing the finite sample performance of the refined estimator to the original gamma kernel approaches, as well as standard boundary-corrected and adjusted fixed kernels. Our results indicate that, in all settings considered, the suggested new estimator offers a superior precision. Finally, we underline the practical relevance of the proposed methodology in two empirical applications, which focus on high-frequency trading volumes and realized volatility estimates, respectively.

Chapter 3 is joint work with my supervisor Nikolaus Hautsch and my former colleague Lada Kyj. A version of this chapter is under revision for publication in the *Journal of Applied Econometrics*. The chapter addresses the unsolved debate about the usefulness of high-frequency data for portfolio allocation given a vast-dimensional asset universe. For that purpose, we consider the general application of selecting global minimum variance portfolios from the constituents of the S&P 500 index over a four-year period, which covers the 2008 financial crisis. Our approach for computing covariance matrix predictions based on high-frequency data relies on a blocked realized kernel estimator, as well as on applying different smoothing windows, various regularization techniques and two forecasting models. In an extensive empirical study, we show that covariance predictions based on high-frequency data yield a significantly lower portfolio volatility when compared to a comprehensive set of benchmark methods employing daily returns. Particularly during the volatile crisis period, these performance gains hold over longer horizons than previous studies have shown and translate into substantial utility benefits from the perspective of an investor exhibiting pronounced risk aversion.

Chapter 1

Capturing the Zero: A New Class of Zero-Augmented Distributions and Multiplicative Error Processes

This chapter is based on Hautsch, Malec, and Schienle (2013).

1.1 Introduction

The availability and increasing importance of high-frequency data in empirical finance and financial practice has triggered the development of new types of econometric models capturing the specific properties of these observations. Typical features of financial data observed on high frequencies are strong serial dependencies, irregular spacing in time, price discreteness and the nonnegativity of various (trading) variables. To account for these properties, models have been developed which contain features of both time series approaches and microeconometric specifications, see, e.g., Engle and Russell (1998), Russell and Engle (2005) or Rydberg and Shephard (2003), among others.

This chapter proposes a novel type of model capturing a further important property of high-frequency data that is present in many situations but not taken into account in extant approaches: the occurrence of a nontrivial part of zeros in the data – henceforth referred to as "excess zeros" – which is a typical phenomenon particularly in the context of high-frequency time aggregates (e.g., 15-second or 30-second data). In high-frequency trading, this type of data is widely used and generally preferred to tick-by-tick level data, as it dispenses with certain pitfalls in econometric modeling, such as the irregular spacing of time spells. However, measures of trading activity within short intervals, such as cumulated trading volumes, naturally reveal a high proportion of zero observations. This is even true for liquid stocks, since there is always a significant proportion of

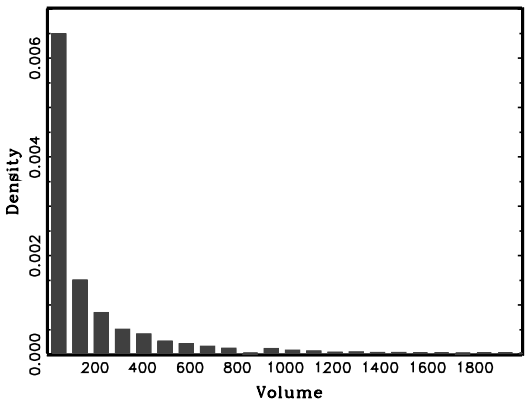


Figure 1.1: Histogram of 15-Second Cumulated Volumes of the McDermott Stock Traded at NYSE for July 2009

intervals with no trading. It should be stressed that such zero clustering effects will also not be mitigated by a further increase of market liquidity over time, as in that case, correspondingly higher frequencies of trading decisions naturally result in smaller aggregation intervals. As a representative illustration, Figure 1.1 depicts the empirical distribution of cumulated trading volumes per 15 seconds of the McDermott stock traded at the New York Stock Exchange (NYSE). No-trade intervals amount to a proportion of about 50%, leading to a significant spike at the leftmost bin.

The occurrence of such high proportions of zero observations can not be appropriately captured by any standard distribution for nonnegative random variables, such as the exponential distribution, generalizations thereof, as well as various types of truncated models (c.f. Johnson et al., 1994). This has serious consequences in a dynamic framework, as e.g., in the multiplicative error model (MEM) introduced by Engle (2002b), which is commonly used to model positive-valued autocorrelated data. In such a framework, employing distributions which do not explicitly account for excess zeros induces severe distributional misspecifications, causing inefficiency and in many cases even inconsistency of parameter estimates. These misspecifications become even more evident when zero occurrences – and thus (no-)trading probabilities – follow their own dynamics. Moreover, standard distributions are clearly inappropriate whenever density forecasts are in the core of interest, since they are not able to explicitly predict zero outcomes.

To the best of our knowledge, existent literature does not provide any systematic and self-contained framework to model, test and predict serially dependent positive-valued data realizing a nontrivial part of excess zeros. Therefore, our main contributions can be summarized as follows. First, we introduce a new type of discrete-continuous mixture distribution capturing a clustering of observations at zero. The idea is to decompose the distribution into a point-mass at zero and a flexible continuous distribution for strictly positive values. Second, we propose a novel semiparametric density test, which is tailored to distributions based on point-mass mixtures. Third, we employ the above mixture

distribution to specify a so-called zero-augmented MEM (ZA-MEM) that allows for maximum likelihood estimation in the presence of zero observations. Finally, we explicitly account for serial dependencies in zero occurrences by introducing an augmented MEM structure which captures the probability of zeros based on a dynamic binary choice component. The resulting so-called Dynamic ZA-MEM (DZA-MEM) yields a specification which allows to explicitly predict zero outcomes and thus is able to produce appropriate density forecasts.

A zero-augmented model is an important complement to current approaches, which reveal clear deficiencies and weaknesses in the presence of zeros. Many distributions for positive-valued random variables, such as the Weibull or gamma distribution and generalizations thereof, imply log-likelihood functions which cannot be evaluated in the case of zero observations. The same is true for a log-normal distribution yielding consistency in a QML setting for a logarithmic MEM (Allen et al., 2008). An exception is the exponential distribution, which allows for positive- *and* zero-valued random variables. In fact, the latter is the only distribution allowing for (consistent) QML estimation of MEMs, while still implying a tractable log-likelihood function in the presence of zeros. Although exponential QML implies consistency of conditional mean parameters, estimates become quite inefficient in the presence of a high proportion of zeros, as the continuous nature of the exponential distribution causes a severe misspecification at the lower boundary of the support. For an illustration, see Figure 1.1. A similar argument applies if the model is estimated by the generalized method of moments (GMM), which is an alternative way for the consistent estimation of the conditional mean in the presence of zeros (see Brownlees et al., 2010). However, the inefficiency of estimates based on QML/GMM can be harmful if the sample size is not too high (e.g., induced by local rolling window estimation, see Härdle et al., 2012) and/or if time-aggregated data is sampled at high frequencies, inducing a high proportion of zeros. In these situations, it becomes essential to explicitly capture the point-mass at zero. The latter is even more relevant when researchers are particularly interested in predicting zero realizations and, in addition, when zero occurrences might follow their own dynamics.

Finally, from an economic viewpoint, no-trade intervals provide self-contained information. E.g., in the asymmetric information-based market microstructure model by Easley and O'Hara (1992), the absence of a trade indicates lacking information in the market. Indeed, the question whether to trade and (if yes) how much to trade are separate decisions which do not necessarily imply that no-trade intervals can be considered as the extreme case of low trading volumes. Consequently, the binary process of non-trading might follow its own dynamics other than that of (nonzero) volumes.

This chapter contributes to several strings of literature. First, it adds to the literature on point-mass mixture distributions. An important distinguishing feature of the existing specifications is whether the point-mass at zero is held constant (e.g., Weglarczyk et al., 2005) or explained by a standard (static) binary choice model (e.g., Duan et al., 1983). We extend these approaches by allowing for a dynamic model for zero occurrences. In an MEM context, De Luca and Gallo (2004) or Lanne (2006) employ mixtures of continuous distributions which are typically motivated by economic arguments, such as

trader heterogeneity. The idea of employing a point-mass mixture distribution to model zero values is only mentioned, but not applied, by Cipollini et al. (2006).

Second, our semiparametric specification test contributes to the class of kernel-based specification tests, as e.g., proposed by Fan (1994), Fernandes and Grammig (2005) or Hagmann and Scaillet (2007). None of the existing methods, however, is suitable for distributions including a point-mass component. If applied to MEM residuals, our approach also complements the literature on diagnostic tests for MEM specifications.

Third, since the proposed dynamic zero-augmented MEM comprises a MEM and a dynamic binary choice part, we also extend the literature on component models for high-frequency data, as e.g., in Rydberg and Shephard (2003) or Liesenfeld et al. (2006), among others. While the latter focus on transaction price changes, our model is applicable to various transaction characteristics, as it decomposes a (nonnegative) persistent process into the dynamics of zero values and strictly positive realizations. For instance, the approach can explain the trading probability in a first stage and, given that a trade has occurred, models the corresponding cumulated volume.

We illustrate the usefulness of the proposed modeling and evaluation framework in two steps. First, a simulation study shows the efficiency gains of a ZA-MEM compared to standard models ignoring zero effects and demonstrates the excellent power of the new semiparametric specification test. Second, we apply our methodology to 15-second cumulative volumes of two liquid and two illiquid stocks traded at the NYSE. The resulting sample is exemplary for situations where the amount of excess zeros is not negligible. Using the developed specification test, we show that the ZA-MEM captures the distributional properties of the data very well. Moreover, a density forecast analysis shows that the novel type of MEM structure is successful in explaining the dynamics of zero values and appropriately predicting the entire distribution. The best performance is shown for a DZA-MEM specification where the zero outcomes are modeled using an autoregressive conditional multinomial (ACM) model as proposed by Russell and Engle (2005). In fact, we observe that trading probabilities are quite persistent, while following their own dynamics. Our results demonstrate that the proposed model can serve as a workhorse for modeling and predicting various high-frequency variables and can be extended in different directions.

The remainder of this chapter is structured as follows. In Section 1.2, we introduce a novel point-mass mixture distribution and develop a corresponding semiparametric specification test, which is applied to evaluate the goodness-of-fit based on MEM residuals. Section 1.3 presents the dynamic zero-augmented MEM capturing serial dependencies in zero occurrences. We evaluate the extended model by examining out-of-sample forecasts of conditional densities. Finally, Section 1.4 concludes.

1.2 A Discrete-Continuous Mixture Distribution

1.2.1 Data and Motivation

We analyze high-frequency trading volume data for the four stocks Bank of America (BAC), International Business Machines (IBM), McDermott International (MDR) and Cimarex Energy (XEC), which are traded at the New York Stock Exchange. The first two represent liquid stocks, while the latter two are less liquid as measured by the total share volume in July 2009. The transaction data is extracted from the Trade and Quote (TAQ) database released by the NYSE and covers the trading week from July 27 to 31, 2009. We filter the raw data by deleting transactions that occurred outside regular trading hours from 9:30 am to 4:00 pm. The tick-by-tick data is aggregated by computing cumulated trading volumes over 15-second intervals, resulting in 7795 observations for the four stocks. Modeling and forecasting cumulated volumes on high frequencies is, for instance, crucial for algorithmic trading strategies (see, e.g., Brownlees et al., 2010). To account for the well-known intraday seasonalities (see, e.g., Hautsch, 2004, for an overview), we divide the cumulated volumes by a seasonality component, which is pre-estimated employing a cubic spline function.

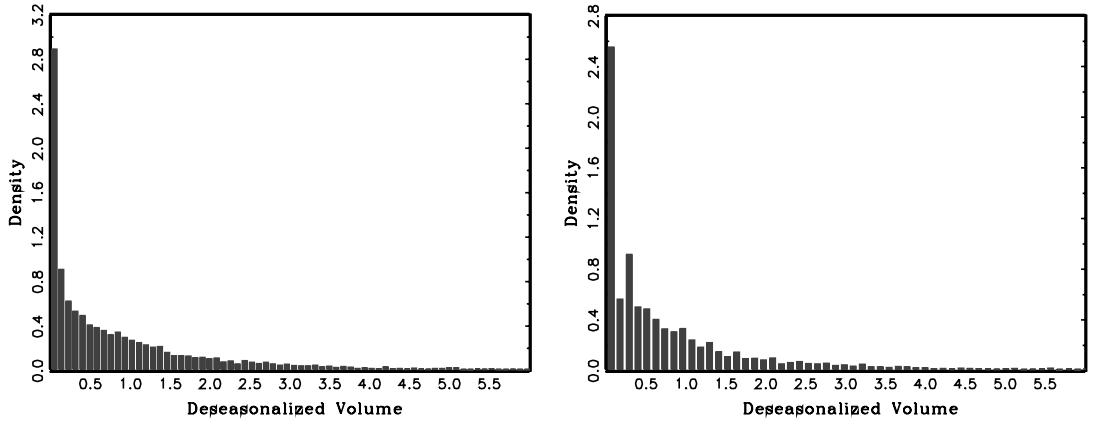
An important feature of the data is the high number of zeros induced by non-trading intervals. The summary statistics in Table 1.1 and the histograms depicted in Figure 1.2 report a nontrivial share of zero observations, ranging from about 9% for BAC to almost 60% for MDR. The proportion of zeros is comparably high, as it is a relatively calm market period. However, we choose this period as an exemplary sample for situations where zeros are not negligible. The latter occur whenever researchers aim at linking the sampling frequency to the underlying (average) trading frequency. Then, more liquid stocks inducing a higher trading intensity also require a higher sampling frequency in order to limit the loss of information on intraday variation. For instance, analyzing the same stocks not in July, but, e.g., in February 2009, would result in a lower proportion of zeros induced by higher trading frequencies. In this case, the same distributional pattern emerges if the sampling frequency is approximately doubled. Likewise, even higher proportions of zeros might be observed if the sampling frequency is further increased or less liquid stocks are analyzed. Therefore, we see the data employed in this chapter as being representative for situations where the necessity of a high sampling frequency or the illiquidity of the underlying assets confronts researchers with a significant proportion of zeros.[1]

A further major feature of cumulated volumes is their strong autocorrelation and high persistence as documented by the Q-statistics in Table 1.1 and the autocorrelation functions (ACFs) displayed in Figure 1.3.

[1]The proportion of zeros is also affected by institutional and technical factors. O'Hara and Ye (2011) show that more than 50% of the trading volume of NYSE stocks is executed on other venues. O'Hara et al. (2011) investigate the fact that the TAQ database contains only transactions with a size of at least 100 shares, although smaller trades can account for up to 66% of the total volume. However, a closer examination of these issues in the given modeling

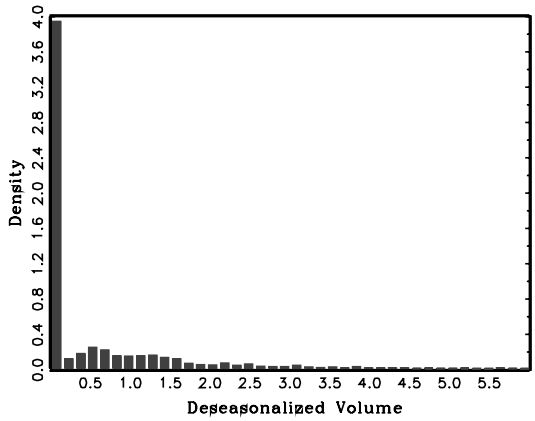
Table 1.1: Summary Statistics of Cumulated Trading Volumes
All statistics are reported for the raw and seasonally adjusted time series. SD: standard deviation, q_5 and q_{95}: 5% and 95% quantile, respectively. n_z/n: share of zero observations. $Q(l)$: Ljung-Box statistic associated with l lags. The 5% (1%) critical values associated with lag lengths 20, 50 and 100 are 31.41 (37.57), 67.51 (76.15) and 124.34 (135.81), respectively.

	BAC		**IBM**	
	Raw	Adj.	Raw	Adj.
Obs	7795	7795	7795	7795
Mean	16612.7	1.02	651.8	1.01
SD	31384.9	1.57	1381.6	1.91
q_5	0	0.00	0	0.00
q_{95}	61800	3.76	2800	3.85
n_z/n	0.092	0.092	0.263	0.263
$Q(20)$	10714.52	1658.86	8614.27	864.52
$Q(50)$	17681.48	2160.17	14431.50	1310.42
$Q(100)$	24795.88	2477.98	17773.13	1575.16
	MDR		**XEC**	
	Raw	Adj.	Raw	Adj.
Obs	7795	7795	7795	7795
Mean	215.2	1.01	163.45	1.01
SD	683.1	3.34	440.88	2.13
q_5	0	0	0	0
q_{95}	900	4.44	700	4.26
n_z/n	0.582	0.582	0.506	0.506
$Q(20)$	3277.38	384.73	3008.03	1118.00
$Q(50)$	4769.92	576.79	4893.87	1615.18
$Q(100)$	6002.79	637.50	5947.61	1891.64

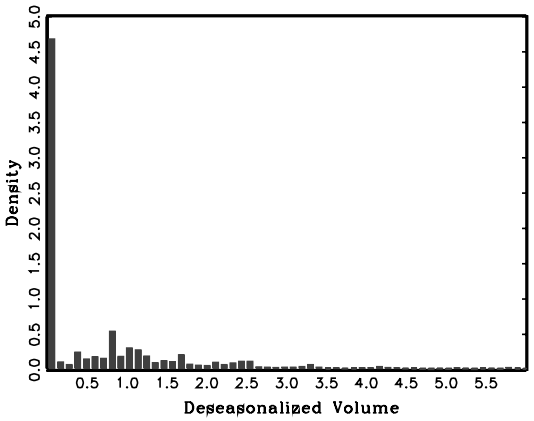


(a) BAC

(b) IBM

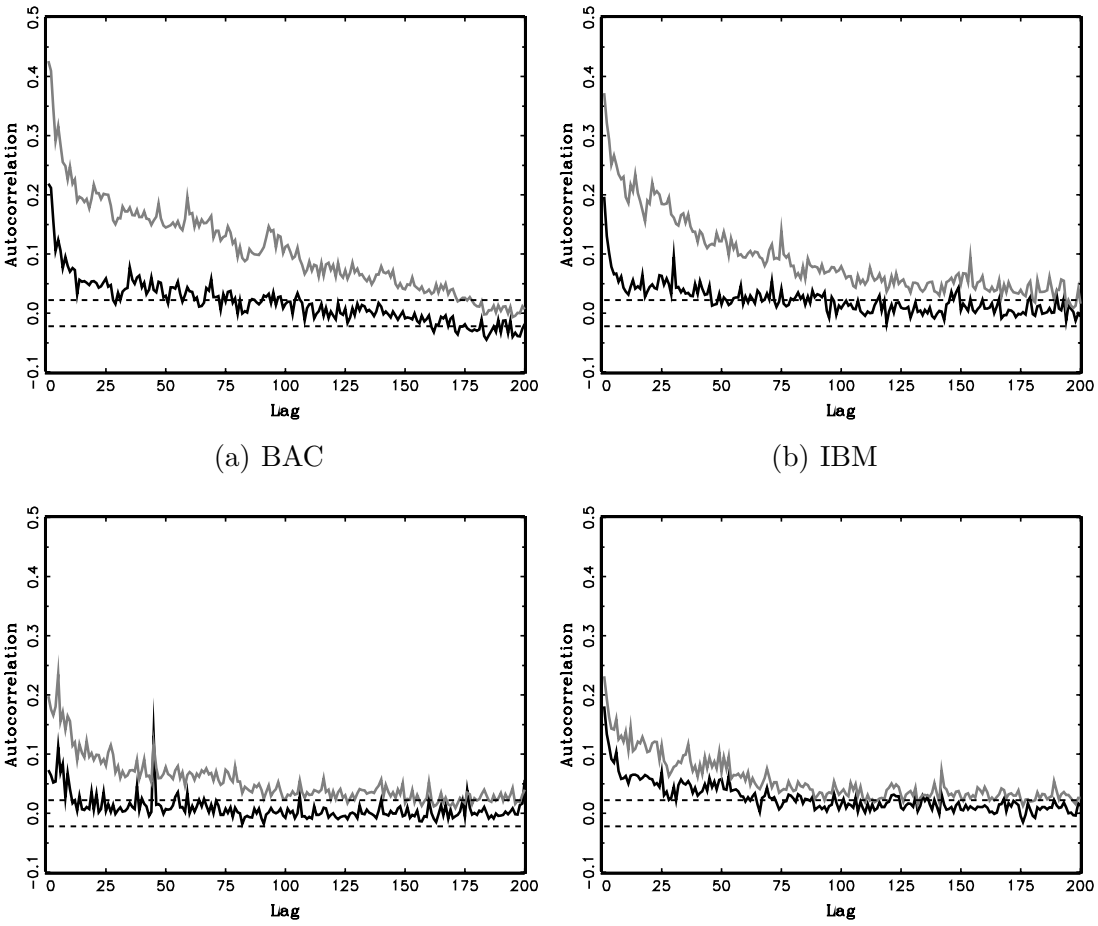


(c) MDR



(d) XEC

Figure 1.2: Sample Histograms of Deseasonalized Cumulated Volumes



(a) BAC (b) IBM

(c) MDR (d) XEC

Figure 1.3: Sample Autocorrelograms
Sample autocorrelation functions of raw (gray line) and diurnally adjusted (black line) cumulated trading volumes. Horizontal lines indicate the limits of 95% confidence intervals ($\pm 1.96/\sqrt{n}$).

To account for these strong empirical features, we first propose a distribution capturing the phenomenon of excess zeros and, secondly, implement it in a MEM setting.

1.2.2 A Zero-Augmented Distribution for Nonnegative Variables

We consider a nonnegative random variable X with independent observations $\{X_t\}_{t=1}^n$, corresponding, e.g., to the residuals of an estimated time series model. In the presence of zero observations, a natural choice is the exponential distribution, as it is also defined for zero outcomes and, being a member of the standard gamma family, provides consistent QML estimates of the underlying conditional mean function (e.g., specified as a MEM). However, in case of high proportions of zero realizations (as documented in Section 1.2.1), this distribution is severely misspecified, making QML estimation quite inefficient.

To account for excess zeros, we assign a discrete probability mass to the exact zero value. Hence, similar to the structure of a tobit, we define the probabilities

$$\pi := P(X > 0), \qquad 1 - \pi := P(X = 0). \tag{1.1}$$

Conditional on $X > 0$, the random variable X follows a continuous distribution with density $g_X(x) := f_X(x|X > 0)$, which is continuous for $x \in (0, \infty)$. Consequently, the unconditional distribution of X is semicontinuous with a discontinuity at zero, implying the density

$$f_X(x) = (1 - \pi)\,\delta(x) + \pi\; g_X(x)\; \mathbb{1}_{(x>0)}, \tag{1.2}$$

where $0 \leq \pi \leq 1$, $\delta(x)$ is a point probability mass at $x = 0$, while $\mathbb{1}_{(x>0)}$ denotes an indicator function taking the value 1 for $x > 0$ and 0 else. The probability π is treated as a parameter of the distribution determining how much probability mass is assigned to the strictly positive part of the support. Note that the above point-mass mixture assumes zero values to be "true" zeros, i.e., they originate from another source than the continuous component and do not result from censoring. This assumption is valid, e.g., in case of cumulative trading volumes, where zero values correspond to non-trade intervals and originate from the decision on whether to trade or not.

The log-likelihood function implied by the mixture density (1.2) is

$$\mathcal{L}(\vartheta) = n_z \ln(1 - \pi) + n_{nz} \ln \pi + \sum_{t \in \mathfrak{I}_{nz}} \ln g_X(x_t; \vartheta_g), \tag{1.3}$$

where $\vartheta = (\pi, \vartheta_g)'$, ϑ_g denotes the vector of parameters determining $g_X(x)$, $\mathfrak{I}_{nz}$ indicates the set of all subscripts t associated with nonzero observations x_t, while n_z and n_{nz} are the number of zero and nonzero observations, respectively. If no dependencies between π and ϑ_g are introduced, componentwise estimation is possible and the estimate of π is given by the empirical frequency of zero observations.

framework goes beyond the scope of this chapter.

The conditional density $g_X(x)$ can be specified according to any distribution defined on positive support. We consider the generalized F (GF) distribution, since it nests most of the distributions frequently used in high-frequency applications (see, e.g., Hautsch, 2003). The corresponding conditional density is given by

$$g_X(x) = \frac{a\, x^{a\,m-1}\, [\eta + (x/\lambda)^a]^{(-\eta-m)}\, \eta^\eta}{\lambda^{a\,m}\, \mathcal{B}(m,\eta)}, \tag{1.4}$$

where $a > 0, m > 0, \eta > 0$ and $\lambda > 0$. $\mathcal{B}(\cdot)$ describes the full beta function with $\mathcal{B}(m,\eta) := \Gamma(m)\,\Gamma(\eta)\,\Gamma(m+\eta)^{-1}$. The conditional noncentral moments implied by the GF distribution are

$$E[X^s|X>0] = \lambda^s\, \eta^{s/a} \frac{\Gamma(m+s/a)\,\Gamma(\eta-s/a)}{\Gamma(m)\,\Gamma(\eta)}, \quad a\,\eta > s. \tag{1.5}$$

Accordingly, the distribution is based on three shape parameters a, m and η, as well as a scale parameter λ. The support of the GF distribution includes the exact zero only if the parameters satisfy the condition $a\,m \geq 1$ with the limiting case of an exponential distribution. A detailed discussion of special cases and density shapes implied by different parameter values can be found, e.g., in Lancaster (1997).

The unconditional density of the zero-augmented generalized F (ZAF) distribution follows from (1.2) and (1.4) as

$$f_X(x) = (1-\pi)\,\delta(x) + \pi\, \frac{a\, x^{a\,m-1}\, [\eta + (x/\lambda)^a]^{(-\eta-m)}\, \eta^\eta}{\lambda^{a\,m}\, \mathcal{B}(m,\eta)}\, \mathbb{1}_{(x>0)}, \tag{1.6}$$

which reduces to the GF density for $\pi = 1$. The unconditional moments can be obtained by exploiting (1.5), i.e.,

$$\begin{aligned} E[X^s] &= \pi\; E[X^s|X>0] + (1-\pi)\;\; E[X^s|X=0]\,, \\ &= \pi\; \lambda^s\, \eta^{s/a} \frac{\Gamma(m+s/a)\,\Gamma(\eta-s/a)}{\Gamma(m)\,\Gamma(\eta)}, \quad a\,\eta > s. \end{aligned} \tag{1.7}$$

The log-likelihood function of the ZAF distribution is given by

$$\begin{aligned} \mathcal{L}\,(\vartheta) = n_z \ln(1-\pi) + n_{nz} \ln \pi + \sum_{t \in \mathfrak{I}_{nz}} \Big\{ & \ln a + (am-1)\; \ln x_t + \eta\; \ln \eta \\ & - (\eta+m)\; \ln\left\{\eta + \left[x_t\, \lambda^{-1}\right]^a\right\} - \ln \mathcal{B}(m,\eta) - am \ln \lambda \Big\}, \end{aligned} \tag{1.8}$$

where $\vartheta = (\pi, a, m, \eta, \lambda)'$.

1.2.3 A New Semiparametric Specification Test

To perform model diagnostics, we introduce a specification test that is tailored to point-mass mixture distributions on nonnegative support like (1.2). Instead of, e.g., checking a number of moment conditions, we consider a kernel-based semiparametric approach, which allows to formally examine whether the entire distribution is correctly specified. Compared to similar smoothing specification tests for densities with left-bounded support, as e.g., proposed by Fernandes and Grammig (2005) and Hagmann and Scaillet (2007), the assumption of a point-mass mixture under the null and alternative hypothesis is a novelty. Estimation in our procedure is optimized for densities which are locally concave for small positive values as described in Section 1.2.1.

In this setting, an appropriate semiparametric benchmark estimator for the unconditional density $f_X(x)$ must have the point-mass mixture structure as in (1.2). Since the support of the discrete and continuous component is disjoint, we can estimate both parts separately without further functional form assumptions. In particular, we use the empirical frequency $\hat{\pi} := n^{-1} \sum_t \mathbb{1}_{(x_t>0)}$ as an estimate for the probability of $X > 0$. The conditional density g_X is estimated using a nonparametric kernel smoother, i.e.,

$$\hat{g}_X(x) = \frac{1}{n_{nz} b} \sum_{t \in \mathfrak{I}_{nz}} K_{x,b}(X_t)\,, \tag{1.9}$$

where K is a kernel function integrating to unity. The estimator is generally consistent on unbounded support for bandwidth choices $b = O(n^{-\nu})$ with $\nu < 1$. However, if the support of the density is bounded, as in our case from below at zero, standard fixed kernel estimators assign weight outside the support at points close to zero, yielding inconsistent results at points near the boundary. Thus instead, we consider a gamma kernel estimator as proposed in Chen (2000) whose flexible form ensures that it is free of boundary bias, while density estimates are always nonnegative. This is in contrast to boundary correction methods for fixed kernels, such as boundary kernels (Jones, 1993) or local-linear estimation (Cheng et al., 1997). The asymmetric gamma kernel is defined on the positive real line and is based on the density of the gamma distribution with shape parameter $x/b + 1$ and scale parameter b, such that

$$K^{\gamma}_{x/b+1,b}(u) := \frac{u^{x/b} \exp(-u/b)}{b^{x/b}\, \Gamma(x/b + 1)}. \tag{1.10}$$

For the final standard gamma kernel estimator, set $K_{x,b}(X_t) = K^{\gamma}_{x/b+1,b}(X_t)$ in (1.9). Note that if the true underlying density has a large probability mass near zero as in our data, it is statistically favorable to employ the standard gamma kernel (1.10), and not the modified version as proposed in Chen (2000) or other boundary correction techniques, such as reflection methods (e.g. Schuster, 1958) or cut-and-normalized kernels (Gasser and Müller, 1979). In this case, the first derivatives of the density are usually significantly nonzero at points close to the boundary. Further, comparing the absolute size of the respective leading terms in the asymptotically vanishing bias expressions of the standard and modified gamma kernel estimator reveals that the sum of first and second derivatives

with opposed signs for the standard gamma kernel estimator is smaller than the pure second derivative for the modified one and the other estimators (see Zhang, 2010, for details). This performance difference is even more relevant in finite samples as outlined in Chapter 2. Note, however, that if, in contrast to our data here, densities were locally convex with no pole at zero, such as for income distributions (see, e.g., Hagmann and Scaillet, 2007), the modified instead of the standard gamma kernel should be used, following exactly the opposite arguments as above.

While for estimation at points further away from the boundary the variance of gamma kernel estimators is smaller when compared to symmetric fixed kernels, their finite sample bias is generally larger. Therefore, we apply a semiparametric correction factor technique as in Hjort and Glad (1995) or Hagmann and Scaillet (2007) to enhance the accuracy of the gamma kernel estimator in the interior of the support. This approach is semiparametric in the sense that the unknown density $g_X(x)$ is decomposed as the product of the initial parametric model $g_X\big(x, \vartheta_g\big)$ and a factor $r(x)$ which corrects for the potentially misspecified parametric start. The estimate of the parametric start is given by $g_X\big(x, \widehat{\vartheta_g}\big)$, where $\widehat{\vartheta_g}$ is the maximum likelihood estimator. The correction factor is estimated by kernel smoothing, such that $\hat{r}(x) = n_{nz}^{-1} \sum_{t \in \mathfrak{I}_{nz}} K_{x/b+1,b}(x_t) \,/ g_X\big(X_t, \widehat{\vartheta_g}\big)$. Therefore, the bias-corrected gamma kernel estimator is

$$\tilde{g}_X(x) = \frac{1}{n_{nz} b} \sum_{t \in \mathfrak{I}_{nz}} K^{\gamma}_{x/b+1,b}(X_t) \; \frac{g_X\big(x, \widehat{\vartheta_g}\big)}{g_X\big(X_t, \widehat{\vartheta_g}\big)}, \tag{1.11}$$

which reduces to the uncorrected estimator if the uniform density is chosen as the initial model. Hjort and Glad (1995) show that a corrected kernel estimator yields a smaller bias than its uncorrected counterpart whenever the correction function is less "rough" than the original density. Their proof is valid for fixed symmetric kernels, but the argument also holds true for gamma-type kernels with slightly modified calculations.

The formal test of the parametric model $f_X\big(x, \vartheta\big)$ against the semiparametric alternative $f_X(x)$ measures discrepancies in squared distances integrated over the support. As the discrete parts coincide in both cases, it is based on

$$I := \pi \int_0^{\infty} \big\{ g_X(x) - g_X\big(x, \vartheta_g\big) \big\}^2 \, dx, \tag{1.12}$$

where $g_X(x)$ and $g_X\big(x, \vartheta_g\big)$ denote the general and parametric conditional densities, respectively. The null and alternative hypothesis are

$$H_0 \colon \; P\Big\{ \hat{f}_X(x) = f_X\big(x, \widehat{\vartheta}\big) \Big\} = 1, \quad H_1 \colon \; P\Big\{ \hat{f}_X(x) = f_X\big(x, \widehat{\vartheta}\big) \Big\} < 1, \tag{1.13}$$

where $\hat{f}_X(x)$ and $f_X\big(x, \widehat{\vartheta}\big)$ are the semiparametric and parametric density estimates with respective continuous conditional parts $\tilde{g}_X(x)$ and $g_X\big(x, \widehat{\vartheta_g}\big)$ as in (1.11). The feasible test statistic is given by

$$T_n := n_{nz} \sqrt{b} \; \hat{\pi} \int_0^{\infty} \Big\{ \tilde{g}_X(x) - g_X\big(x, \widehat{\vartheta_g}\big) \Big\}^2 \, dx. \tag{1.14}$$

Asymptotic normality of T_n could be shown using the results of Fernandes and Monteiro (2005). However, it is well-documented that non- and semiparametric tests suffer from size distortions in finite samples (e.g. Fan, 1998). Therefore, we employ a bootstrap procedure as in Fan (1998) to compute size-corrected p-values. This is outlined in detail in the following subsection for a MEM framework.

We choose the bandwidth b according to least-squares cross-validation, which is fully data-driven and automatic. Thus, for the bias-corrected gamma kernel estimator (1.11), the bandwidth b must minimize

$$CV(b) := \frac{1}{n_{nz}^2} \sum_{i \in \mathfrak{I}_{nz}} \sum_{j \in \mathfrak{I}_{nz}} \frac{\int_0^\infty g_X\big(x, \widehat{\vartheta_g}\big)^2 \, K^\gamma_{x/b+1,b}(x_i) \; K^\gamma_{x/b+1,b}(x_j) \; dx}{g_X\big(x_i, \widehat{\vartheta_g}\big) g_X\big(x_j, \widehat{\vartheta_g}\big)}$$
$$- \frac{2}{n_{nz}\,(n_{nz}-1)} \sum_{i \in \mathfrak{I}_{nz}} \sum_{j \neq i \in \mathfrak{I}_{nz}} K^\gamma_{x_i/b+1,b}(x_j) \, \frac{g_X\big(x_i, \widehat{\vartheta_{g_{(i)}}}\big)}{g_X\big(x_j, \widehat{\vartheta_{g_{(i)}}}\big)}, \tag{1.15}$$

where $\widehat{\vartheta_{g_{(i)}}}$ denotes the maximum likelihood estimate computed without observation X_i. The cross-validation objective function is directly derived from requiring the bandwidth to minimize the integrated squared distance between the semiparametric and parametric estimates. For the uncorrected gamma kernel estimator, the corresponding objective function is analogous to (1.15), but does not involve density terms.

Our test differs from related methods not only by being designed for point-mass mixtures. Fan (1994) uses fixed kernels with the respective boundary consistency problems. Fully nonparametric (uncorrected) gamma kernel-based tests as Fernandes and Grammig (2005) have a larger finite sample bias near the boundary for locally concave densities and generally also in the interior of the support. The semiparametric test by Hagmann and Scaillet (2007) suffers from the same problem near zero. Furthermore, weighting with the inverse of the parametric density in their test statistic yields a particularly poor fit in regions with sparse probability, which is an issue in our application, as the distributions are heavily right-skewed.

1.2.4 Empirical and Simulation-Based Evidence for a Zero-Augmented MEM

In order to apply the proposed specification test to our data, we have to appropriately capture the serial dependence in cumulated volumes. This task is performed by specifying a multiplicative error model (MEM) based on a zero-augmented distribution. Accordingly, cumulated volumes y_t are given by

$$y_t = \mu_t \, \varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d. } \mathcal{D}(1)\,, \tag{1.16}$$

where μ_t denotes the conditional mean given the information set $\mathcal{F}_{t-1}$ and depending on a parameter vector ϑ_μ, i.e., $\mu_t := E[y_t|\mathcal{F}_{t-1}] = \mu(\mathcal{F}_{t-1}; \vartheta_\mu)$. ε_t denotes a disturbance following a distribution $\mathcal{D}(1)$ with nonnegative support and $E[\varepsilon_t] = 1$. A deeper discussion of the properties of MEMs is given by Engle (2002b) or Engle and Gallo (2006). We

specify μ_t in terms of a logarithmic specification proposed by Bauwens and Giot (2000) for autoregressive conditional duration (ACD) models which does not require parameter constraints to ensure the positivity of μ_t. Accordingly, μ_t is given by

$$\ln \mu_t = \omega + \sum_{i=1}^{p} \alpha_i \, \ln \varepsilon_{t-i} \, \mathbb{1}_{(y_{t-i}>0)} + \sum_{i=1}^{p} \alpha_i^0 \, \mathbb{1}_{(y_{t-i}=0)} + \sum_{i=1}^{q} \beta_i \, \ln \mu_{t-i}, \tag{1.17}$$

where the additional dummy variables prevent the computation of $\ln \varepsilon_{t-i}$ whenever $\varepsilon_{t-i} = 0$. The lag structure is chosen according to the Schwarz information criterion (SIC). For more details on the properties of the logarithmic MEM, we refer to Bauwens and Giot (2000) and Bauwens et al. (2003). A survey of additional MEM specifications is provided by Bauwens and Hautsch (2008).

Define the zero-augmented MEM (ZA-MEM) as a MEM where ε_t is distributed according to the ZAF density (1.6) with scale parameter $\lambda = (\pi \, \xi)^{-1}$ and

$$\xi := \eta^{1/a} \left[\Gamma(m + 1/a) \, \Gamma(\eta - 1/a)\right] \left[\Gamma(m) \, \Gamma(\eta)\right]^{-1}. \tag{1.18}$$

Recalling (1.7), the constraint on λ ensures that the unit mean assumption for ε_t is fulfilled. The MEM structure (1.16) implies that, conditionally on the information set $\mathcal{F}_{t-1}$, y_t follows a ZAF distribution with $\lambda_t = \mu_t \, (\pi \, \xi)^{-1}$. Note that the latter constraint prevents componentwise optimization of the corresponding log-likelihood and thus requires joint estimation of all parameters.

To implement the semiparametric specification test (1.14) in the above MEM setting, we estimate the model by exponential QML. This approach yields residuals $\hat{\varepsilon}_t := y_t / \hat{\mu}_t$, which are consistent estimates of the i.i.d. errors ε_t. Alternatively, we could obtain consistent error estimates using the semiparametric methods by Drost and Werker (2004) or employing GMM as in Brownlees et al. (2010). The consistency and parametric rate of convergence of the conditional mean estimates enable us to use the residuals as inputs for the semiparametric specification test without affecting the asymptotics of the kernel estimators discussed in Section 1.2.3. A similar procedure is applied by Fernandes and Grammig (2005) for their nonparametric specification test. Finally, we obtain applicable finite sample p-values by employing the following bootstrap procedure:

Step 1: Draw a random sample $\{\varepsilon_t^*\}_{t=1}^n$ from the parametric ZAF distribution with density $f_\varepsilon\big(\varepsilon, \widehat{\vartheta}\big)$, where $\widehat{\vartheta}$ is the maximum likelihood estimate of the ZAF parameters ϑ based on the original data. From this, generate a bootstrap sample $\{y_t^*\}_{t=1}^n$ as $y_t^* = \hat{\mu}_t \, \varepsilon_t^*$, where $\hat{\mu}_t$ is the fitted conditional mean as in (1.17) based on the maximum likelihood estimates from the original data.
Step 2: Use $\{y_t^*\}_{t=1}^n$ to compute the statistic T_n, which we denote as T_n^*. This requires the re-evaluation of both the parametric and semiparametric estimates of $f_\varepsilon(\varepsilon)$.
Step 3: Steps 1 and 2 are repeated B times and p-values are obtained from the empirical distribution of $\big\{T_{n,r}^*\big\}_{r=1}^B$.

Before the empirical application, we conduct a simulation study to investigate the following two issues: the inefficiency of parameter estimates based on an error distribution that does not capture zero clustering effects, as well as the power of the proposed

specification test in a MEM setting. We consider four data-generating processes (DGPs) which assume the above zero-augmented MEM structure relying on the ZAF distribution as in (1.6), (1.16), (1.17) and (1.18) with parameter values chosen to replicate the stylized facts of the data. For each DGP, 1000 samples with 8000 observations are simulated.

To address the first question, we estimate the MEM parameters by maximum likelihood based on the ZAF distribution and by QML based on the (misspecified) exponential distribution. Table 1.2 displays the simulation results for the different scenarios. Despite the considerable sample size, the ML estimates of the ZA-MEM consistently exhibit lower standard deviations and root mean squared errors (RMSEs). The discrepancy in precision is more pronounced for DGPs with a larger value of the shape parameter m of the ZAF distribution and a higher probability of zero outcomes. The latter finding demonstrates the relationship between the magnitude of zero clustering and the relative inefficiency of the exponential QML approach compared to the ML estimator of the ZA-MEM.

For the power study, we estimate three models. All assume the MEM structure (1.16) with a correctly specified conditional mean μ_t and errors ε_t following the general zero-augmented distribution in (1.1) and (1.2). However, they introduce different misspecifications of the conditional error density $g_\varepsilon(\varepsilon_t)$. The first model (E-ZA-MEM) assumes an exponential distribution with scale parameter $\lambda = \pi^{-1}$, while the second one (G-ZA-MEM) considers a gamma distribution with shape parameter m and scale parameter $\lambda = (\pi\, m)^{-1}$. The third specification (W-ZA-MEM) assumes a Weibull distribution with shape parameter a and scale parameter $\lambda = (\pi\, \xi_w)^{-1}$, where $\xi_w := \Gamma(1 + 1/a)$. Table 1.3 displays the rejection rates of the specification test based on 500 bootstrap replications for the p-values. For all DGPs and (misspecified) models, the rejection rates are close or equal to one. Accordingly, the proposed test exhibits a high power regarding the detection of misspecified error distributions in various scenarios, which indicates that it constitutes a reliable inference technique in empirical applications.

We now apply the above estimation and testing methodology to the cumulated volume data. Table 1.4 shows the maximum likelihood estimates of the ZA-MEM based on the ZAF distribution, while Figure 1.4 depicts the resulting parametric error densities together with their semiparametric counterparts based on the uncorrected gamma kernel. For all stocks, the parametric and semiparametric densities are quite close to each other. However, there is a noticeable discrepancy to the right of the boundary, which can be explained by the increased bias of the gamma kernel compared to standard fixed kernels in the interior of the support. To refine the semiparametric density estimate, we employ the bias-corrected gamma kernel estimator (1.11), choosing the ZAF distribution as parametric start. The plots in Figure 1.5 show that, in all cases, the discrepancy between both estimates vanishes, as the parametric density now generally lies within the 95% confidence region of the semiparametric estimate. For the less liquid stocks MDR and XEC, the density estimates are virtually zero on an interval near the lower boundary of the support. Since the parametric density serves as the starting model for the corrected gamma kernel estimator, the vanishing probability mass close to the origin also explains the large cross-validation bandwidths. More details on the relationship

Table 1.2: Simulation Results – ZA-MEM vs. Exponential QML
Each DGP assumes a zero-augmented Log-MEM based on the ZAF distribution and the MEM parameters $\omega = 0.05$, $\alpha_1 = 0.05$, $\beta_1 = 0.9$ and $\alpha_1^0 = -0.005$. For every replication, MEM parameters are estimated by ML based on the ZAF distribution and by exponential QML. The study uses 1000 replications and a sample size of 8000. SD denotes the standard deviation, RMSE is the root mean squared error.

	ZA-MEM				**Exp. QML**			
	$\hat{\omega}$	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\alpha}_1^0$	$\hat{\omega}$	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\alpha}_1^0$
DGP 1: $a = 0.6$, $m = 100$, $\eta = 3.3$, $\pi = 0.9$								
Median	0.0510	0.0500	0.8990	-0.0048	0.0505	0.0508	0.8962	-0.0078
Mean	0.0512	0.0501	0.8977	-0.0047	0.0630	0.0529	0.8722	-0.0030
SD	0.0082	0.0061	0.0153	0.0169	0.0586	0.0220	0.1165	0.0697
RMSE	0.0082	0.0061	0.0154	0.0169	0.0600	0.0221	0.1198	0.0697
DGP 2: $a = 0.6$, $m = 100$, $\eta = 3.3$, $\pi = 0.5$								
Median	0.0510	0.0505	0.8988	-0.0057	0.0552	0.0535	0.8892	-0.0076
Mean	0.0539	0.0506	0.8946	-0.0058	0.1021	0.0589	0.8155	-0.0052
SD	0.0212	0.0113	0.0327	0.0147	0.1662	0.0453	0.2407	0.0695
RMSE	0.0216	0.0113	0.0331	0.0147	0.1741	0.0462	0.2549	0.0695
DGP 3: $a = 0.6$, $m = 1.9$, $\eta = 100$, $\pi = 0.9$								
Median	0.0504	0.0502	0.8987	-0.0039	0.0503	0.0501	0.8986	-0.0036
Mean	0.0507	0.0501	0.8981	-0.0045	0.0507	0.0503	0.8978	-0.0038
SD	0.0072	0.0057	0.0144	0.0220	0.0077	0.0061	0.0156	0.0231
RMSE	0.0072	0.0057	0.0146	0.0220	0.0077	0.0061	0.0158	0.0232
DGP 4: $a = 0.6$, $m = 1.9$, $\eta = 100$, $\pi = 0.5$								
Median	0.0510	0.0499	0.8980	-0.0054	0.0511	0.0504	0.8970	-0.0038
Mean	0.0538	0.0505	0.8938	-0.0053	0.0552	0.0512	0.8895	-0.0033
SD	0.0210	0.0112	0.0332	0.0190	0.0306	0.0135	0.0498	0.0241
RMSE	0.0213	0.0112	0.0338	0.0190	0.0310	0.0135	0.0508	0.0241

Table 1.3: Simulation Results – Power of Semiparametric Specification Test
Rejection rates of the semiparametric specification test for the distribution of MEM errors ε_t. The same DGPs as in Table 1.2 are used. For every replication, we estimate three models based on the zero-augmented MEM structure with a misspecified distribution of the strictly positive errors. The specification test considers empirical p-values based on 500 bootstrap replications. Following Fernandes and Grammig (2005), rule-of-thumb bandwidths adjusted to gamma kernels and using the exponential distribution as reference are employed, i.e., $\hat{b} = 4^{-1/5}\,\hat{\lambda}\left(\hat{\lambda}-1/2\right)^{-4/5} n_{nz}^{-4/9}$, where $\hat{\lambda}$ is the sample mean of strictly positive observations. The study uses 1000 replications and a sample size of 8000.

	DGP 1		**DGP 2**		**DGP 3**		**DGP 4**	
Est. Model$\backslash\alpha$	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01
E-ZA-MEM	1.000	0.999	1.000	0.999	1.000	1.000	1.000	1.000
G-ZA-MEM	1.000	0.999	0.999	0.999	1.000	1.000	1.000	1.000
W-ZA-MEM	1.000	0.999	0.999	0.999	1.000	1.000	0.996	0.987

between the shape of the parametric start density and the optimal bandwidth can be found in Hjort and Glad (1995).

The estimation results suggest that the ZAF distribution provides a superior way to model MEM disturbances for cumulated volumes. This graphical intuition can be formally assessed by the semiparametric specification test (1.14). Table 1.5 displays the test results based on 1000 bootstrap replications for the empirical p-values. In all four cases, the statistic is insignificant at all conventional levels, which implies that we cannot reject the null hypothesis (1.13). These results confirm that the ZA-MEM is able to capture the distributional properties of high-frequency cumulated volumes.

1.3 Dynamic Zero-Augmented Multiplicative Error Models

1.3.1 Motivation

Assumption (1.1) implies that, conditional on past information, the trading probability is constant or, more formally,

$$\pi := P(\varepsilon_t > 0|\mathcal{F}_{t-1}) = P(y_t > 0|\mathcal{F}_{t-1}) = P(\mathcal{I}_t = 1|\mathcal{F}_{t-1})\,, \tag{1.19}$$

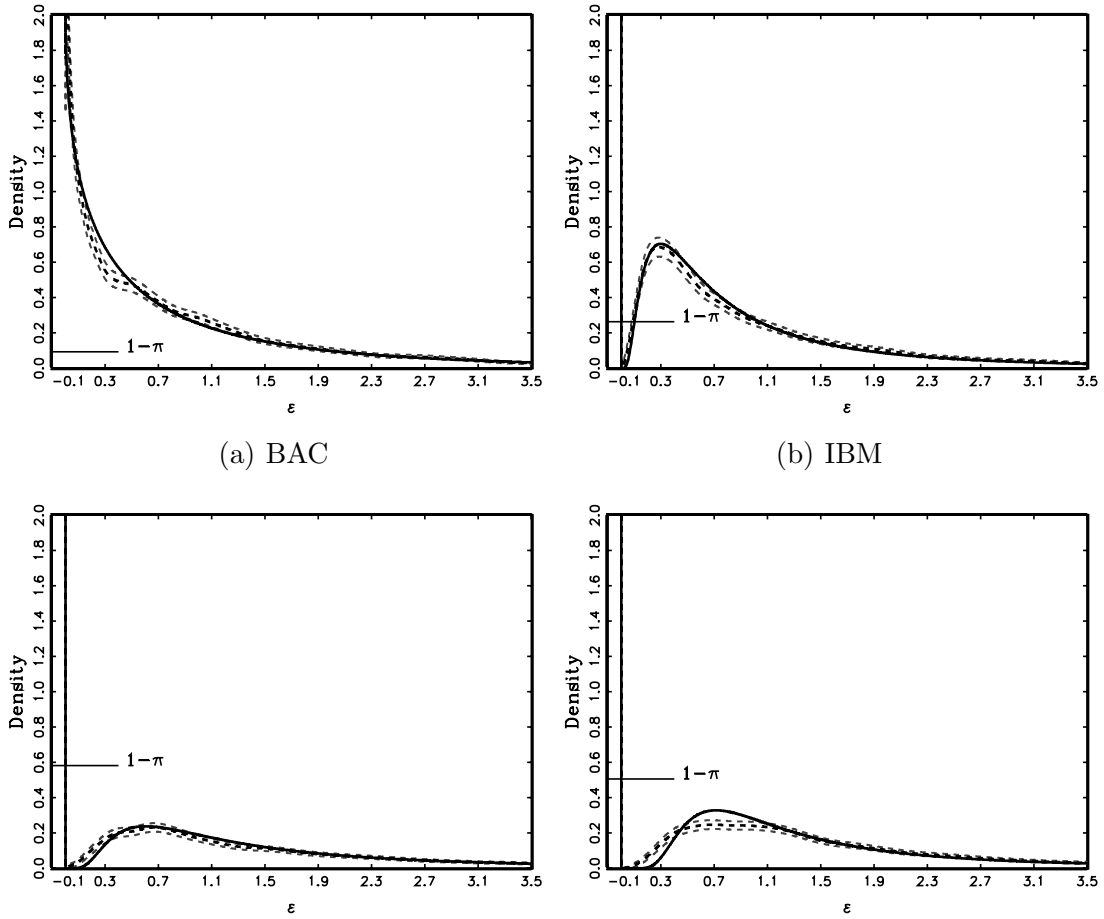
where $\mathcal{I}_t$ is a "trade indicator" taking the value 1 for $y_t > 0$ and 0 else. The assumption of constant no-trade probabilities is in line with the seminal model of nonsynchronous trading by Lo and MacKinlay (1990) but appears to be rather restrictive, as (nonzero) cumulative volume is clearly time-varying and reveals persistent serial dependencies.

Table 1.4: Estimation Results – ZA-MEM
Maximum likelihood estimates and t-statistics of the zero-augmented Log-MEM based on the ZAF distribution. Lag structure is determined using the SIC.

	BAC		**IBM**		**MDR**		**XEC**	
	Coef.	T-St.	Coef.	T-St.	Coef.	T-St.	Coef.	T-St.
ω	0.041	6.301	0.017	6.750	-0.028	-5.379	-0.023	-8.108
α_1	0.118	8.808	0.187	13.757	0.091	9.259	0.130	6.554
α_2	-0.060	-3.731	-0.119	-7.856	-	-	-0.066	-3.181
β_1	0.913	64.253	0.930	135.281	0.938	116.017	0.953	206.477
α_1^0	-0.315	-3.328	-0.162	-5.235	0.032	4.831	-0.013	-0.551
α_2^0	0.291	3.171	0.144	4.640	-	-	0.044	1.926
m	1.703	3.871	653.758	41.981	450.064	8.379	507.419	13.310
η	562.562	12.143	7.533	7.696	3.343	5.335	1.856	14.411
a	0.570	6.748	0.385	14.620	0.642	9.893	1.084	24.059
π	0.908	277.210	0.737	147.718	0.419	74.887	0.495	87.677
$\mathcal{L}$	-9335.306		-10850.092		-10452.980		-10917.378	
SIC	18760.222		21789.796		20977.645		21924.368	

Table 1.5: Semiparametric Specification Test
Results of the semiparametric specification test applied to the MEM errors ε_t. The reported p-values are based on the empirical distribution of the test statistic resulting from 1000 simulated bootstrap samples.

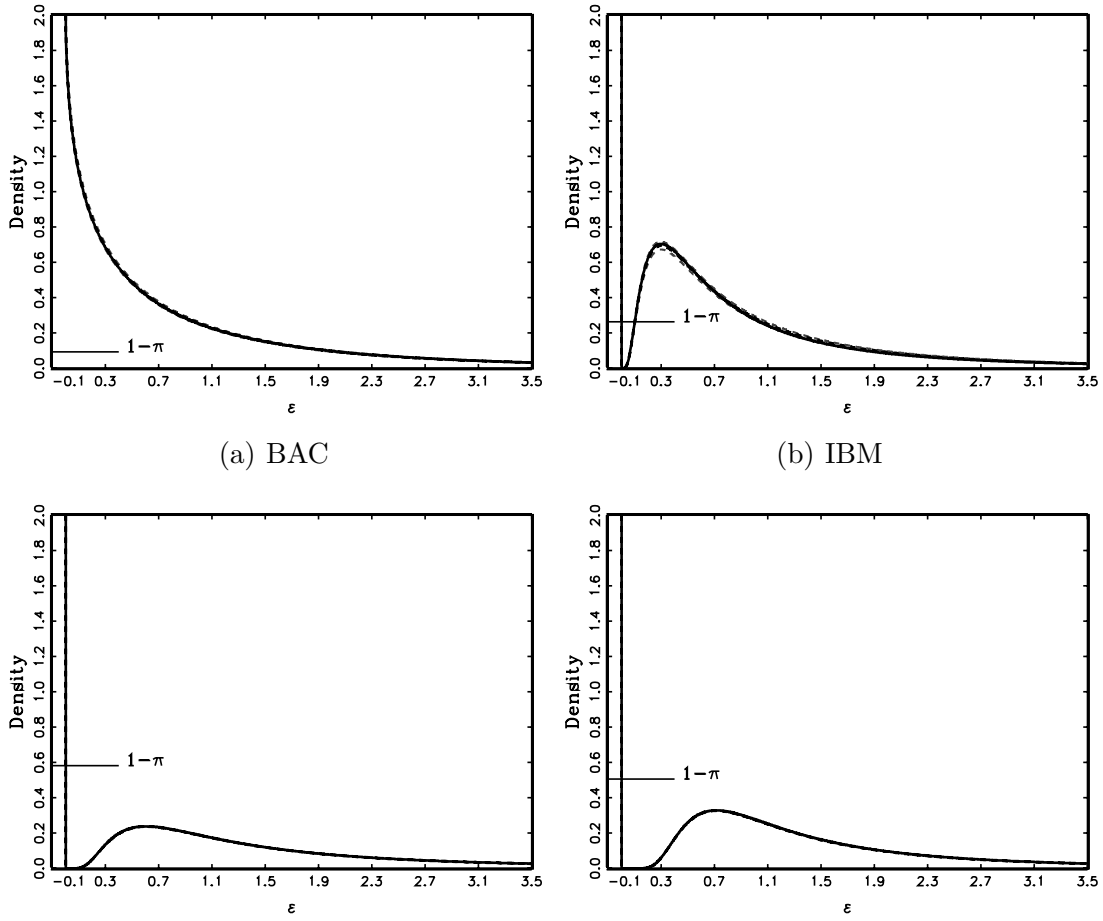
	BAC	**IBM**	**MDR**	**XEC**
T_n	0.298	0.818	1.404	1.308
P-Val.	0.208	0.164	0.990	0.972



(a) BAC (b) IBM

(c) MDR (d) XEC

Figure 1.4: Estimates of Error Density with Gamma KDE
The black solid line represents the error density implied by the ML estimates of the ZA-MEM. The black dashed line is the semiparametric estimate based on the gamma kernel estimator. The gray dashed lines are 95% confidence bounds of the kernel density estimator. CV bandwidths: 0.020 (BAC), 0.012 (IBM), 0.004 (MDR), 0.003 (XEC). Estimates of $1 - \pi$ based on sample percentage of zeros values: 0.092 (BAC), 0.263 (IBM), 0.582 (MDR), 0.506 (XEC).



(a) BAC (b) IBM

(c) MDR (d) XEC

Figure 1.5: Estimates of Error Density with Corrected Gamma KDE
The black solid line represents the error density implied by the ML estimates of the ZA-MEM. The black dashed line is the semiparametric estimate based on the bias-corrected gamma kernel estimator. The gray dashed lines are 95% confidence bounds of the kernel density estimator. CV bandwidths: 1.455 (BAC), 0.406 (IBM), 10578.031 (MDR), 1096.787 (XEC).

Table 1.6: Runs Test for the Trade Indicator
Results of the two-sided runs test for serial dependence of the indicator for nonzero aggregated volumes. Under the null of no serial dependence, the statistic $Z = \frac{R-E(R)}{V(R)}$ (R: number of runs) is asymptotically standard normal.

	BAC	**IBM**	**MDR**	**XEC**
Z	-10.000	-13.832	-17.558	-17.015
P-Val.	0.000	0.000	0.000	0.000

Moreover, it is at odds with the well-known empirical evidence of autocorrelated trading intensities (see, e.g., Engle and Russell, 1998). Table 1.6 shows the results of a simple runs test based on the trade indicator $\mathcal{I}_t$, suggesting that the null hypothesis of no serial correlation in no-trade probabilities is clearly rejected. To capture this effect, we propose an augmented version of the ZA-MEM accounting also for dynamics in zero occurrences.

1.3.2 A ZA-MEM with Dynamic Zero Probabilities

Assume that, given the information set $\mathcal{F}_{t-1}$, the conditional probability of the disturbance ε_t being zero depends on a restricted information set $\mathcal{H}_{t-1} \subset \mathcal{F}_{t-1}$. Moreover, π_t is assumed to depend on $\mathcal{H}_{t-1}$ by a function $\pi(\cdot;\vartheta_\pi)$ with parameter vector ϑ_π, i.e.,

$$\pi_t := P(\varepsilon_t > 0|\mathcal{F}_{t-1}) = P(\varepsilon_t > 0|\mathcal{H}_{t-1}) = \pi(\mathcal{H}_{t-1};\vartheta_\pi)\,. \tag{1.20}$$

As a consequence of this assumption, the disturbances lose the i.i.d. property and, conditionally on $\mathcal{H}_{t-1}$, are independently but *not* identically distributed. Thus, the dynamics of the endogenous variable y_t are not fully captured by the conditional mean μ_t, as past information contained in $\mathcal{H}_{t-1}$ affects the innovation distribution. Similar generalizations of the MEM error structure have been considered, e.g., by Zhang et al. (2001) or Drost and Werker (2004). The resulting dynamic zero-augmented MEM (DZA-MEM) can be formally written as

$$y_t = \mu_t\,\varepsilon_t, \quad \varepsilon_t|\mathcal{H}_{t-1} \sim \text{i.n.i.d. } \mathcal{PMD}(1)\,, \tag{1.21}$$

where $\mathcal{PMD}(1)$ denotes a point-mass mixture as in (1.2) with assumption (1.1) replaced by (1.20) and $E[\varepsilon_t|\mathcal{H}_{t-1}] = E[\varepsilon_t] = 1$. Hence, the conditional density of ε_t given $\mathcal{H}_{t-1}$ is

$$f_\varepsilon(\varepsilon_t|\mathcal{H}_{t-1}) = (1-\pi_t)\,\delta(\varepsilon_t) + \pi_t\; g_\varepsilon(\varepsilon_t|\mathcal{H}_{t-1})\,\mathbb{1}_{(\varepsilon_t>0)}, \tag{1.22}$$

where the conditional density for $\varepsilon_t > 0$, $g_\varepsilon(\varepsilon_t|\mathcal{H}_{t-1})$, depends on $\mathcal{H}_{t-1}$ through the probability π_t, as the unit mean assumption in (1.21) requires

$$\kappa_t := E[\varepsilon_t|\varepsilon_t > 0;\mathcal{H}_{t-1}] = \pi_t^{-1}, \tag{1.23}$$

such that

$$E[\varepsilon_t] = E\{E[\varepsilon_t|\mathcal{H}_{t-1}]\} = E[\pi_t\,\kappa_t] = 1. \tag{1.24}$$

Since the function $\pi(\cdot;\vartheta_\pi)$ is equivalent to a binary choice specification for the trade indicator $\mathcal{I}_t$ defined in (1.19), the log-likelihood of the DZA-MEM consists of a MEM and a binary choice part,

$$\begin{aligned}\mathcal{L}(\vartheta) = &\sum_{t=1}^{n} \{\mathcal{I}_t\,\ln\pi(\mathcal{H}_{t-1};\vartheta_\pi) + (1-\mathcal{I}_t)\,\ln\left[1-\pi(\mathcal{H}_{t-1};\vartheta_\pi)\right]\} \\ &+ \sum_{t\in\mathfrak{I}_{nz}} \{\ln f_\varepsilon(y_t/\mu(\mathcal{F}_{t-1};\vartheta_\mu)\,|\mathcal{H}_{t-1};\vartheta_g) - \ln\mu(\mathcal{F}_{t-1};\vartheta_\mu)\}\,,\end{aligned} \tag{1.25}$$

where $\vartheta = (\vartheta_\pi, \vartheta_g, \vartheta_\mu)'$. As in the previous section, a separate optimization of the two parts is infeasible, since the constraint (1.23) implies that both components depend on the parameters of the binary choice specification, ϑ_π.

If we use the ZAF distribution as point-mass mixture $\mathcal{PMD}(1)$, we obtain the conditional density of ε_t given $\mathcal{H}_{t-1}$ as

$$f_\varepsilon(\varepsilon_t|\mathcal{H}_{t-1}) = (1-\pi_t)\,\delta(\varepsilon_t) + \pi_t\,\frac{a\,\varepsilon_t^{a\,m-1}\left[\eta + (\varepsilon_t\,\pi_t\,\xi)^a\right]^{(-\eta-m)}\,\eta^\eta}{(\pi_t\,\xi)^{-a\,m}\;\mathcal{B}(m,\eta)}\;\mathrm{II}_{(\varepsilon_t>0)}, \tag{1.26}$$

where we set $\lambda_t = (\pi_t\,\xi)^{-1}$, with ξ defined as in (1.18), to meet the constraint (1.23). The corresponding log-likelihood function is

$$\begin{aligned}\mathcal{L}(\vartheta) = &\sum_{t=1}^{n} \{\mathcal{I}_t\,\ln\pi(\mathcal{H}_{t-1};\vartheta_\pi) + (1-\mathcal{I}_t)\,\ln\left[1-\pi(\mathcal{H}_{t-1};\vartheta_\pi)\right]\} \\ &+ \sum_{t\in\mathfrak{I}_{nz}} \left\{ \log a + (am-1)\,\ln y_t - (\eta+m)\,\ln\left\{\eta + \left[y_t\,\frac{\mu(\mathcal{F}_{t-1};\vartheta_\mu)}{\pi(\mathcal{H}_{t-1};\vartheta_\pi)}\,\xi\right]^a\right\}\right. \\ &\left. + \eta\,\ln\eta - am\ln\left[\frac{\mu(\mathcal{F}_{t-1};\vartheta_\mu)}{\pi(\mathcal{H}_{t-1};\vartheta_\pi)}\,\xi^{-1}\right] - \ln\mathcal{B}(m,\eta)\right\},\end{aligned} \tag{1.27}$$

where $\vartheta = (\vartheta_\pi, a, m, \eta, \vartheta_\mu)'$.

1.3.3 Dynamic Models for the Trade Indicator

In order to allow the trade indicator $\mathcal{I}_t$ to follow a dynamic process, we propose two alternative specifications: a parsimonious autologistic specification and a more flexible parameterization using autoregressive conditional multinomial (ACM) dynamics as proposed by Russell and Engle (2005). By considering the general logistic link function

$$\pi_t = \pi(\mathcal{H}_{t-1};\vartheta_\pi) = \frac{\exp(h_t)}{1+\exp(h_t)}, \tag{1.28}$$

the autologistic specification for $h_t = \ln[\pi_t/(1-\pi_t)]$ is given by

$$h_t = \theta_0 + \sum_{i=1}^{l} \theta_i \, \Delta_{t-i} + \sum_{i=1}^{d} \gamma_i \, \mathcal{I}_{t-i}, \tag{1.29}$$

where Δ_t denotes an indicator for large values of the endogenous variable y_t and is defined as

$$\Delta_t := \max(y_t - \mathcal{I}_t, 0)\,. \tag{1.30}$$

This type of transformation was suggested in a similar setting by Rydberg and Shephard (2003) and accounts for the multicollinearity between the lags of y_t and $\mathcal{I}_t$. The autologistic model has advantages in terms of tractability, such as the concavity of the log-likelihood function, making numerical maximization straightforward. However, since this process does not include a moving average component, it is not able to capture persistent dynamics in the binary sequence. Therefore, as an alternative, we propose an ACM specification given by

$$h_t = \varpi + \sum_{j=1}^{v} \rho_j \, s_{t-j} + \sum_{j=1}^{w} \zeta_j \, h_{t-j}, \tag{1.31}$$

where

$$s_{t-j} := \frac{\mathcal{I}_{t-j} - \pi_{t-j}}{\sqrt{\pi_{t-j}\,(1-\pi_{t-j})}}, \tag{1.32}$$

denotes the standardized trade indicator. The process $\{s_t\}$ is a martingale difference sequence with zero mean and unit conditional variance, which implies that $\{h_t\}$ follows an ARMA process driven by a weak white noise term. Consequently, $\{h_t\}$ is stationary if all values of z satisfying $1 - \zeta_1 z - \ldots - \zeta_w z^w = 0$ lie outside the unit circle. For more details, see Russell and Engle (2005).

An appealing feature of the ACM specification in the given framework is its similarity to a MEM. Actually, analogously to a MEM specification, it imposes a linear autoregressive structure on the logistic transformation of the probability π_t, which in turn, equals the conditional mean of the trade indicator $\mathcal{I}_t$ given the restricted information set $\mathcal{H}_{t-1}$, i.e., $E\left[\mathcal{I}_t|\mathcal{H}_{t-1}\right]$.

The DZA-MEM dynamics can be straightforwardly extended by covariates which allow to test specific market microstructure hypotheses. Moreover, a further natural extension of the DZA-MEM is to allow for dynamic interaction effects between the conditional mean of y_t, μ_t, and the probability of zero values, π_t. For instance, by allowing for spillovers between both dynamic equations, the DZA-MEM can be modified as

$$h_t = \varpi + \sum_{j=1}^{v} \rho_j \, s_{t-j} + \sum_{j=1}^{w} \zeta_j \, h_{t-j} + \sum_{j=1}^{m^*} \tau_j \, \mu_{t-j}, \tag{1.33}$$

$$\ln \mu_t = \omega + \sum_{i=1}^{p} \alpha_i \, \ln \varepsilon_{t-i} \, \mathbb{1}_{(y_{t-i}>0)} + \sum_{i=1}^{p} \alpha_i^0 \, \mathbb{1}_{(y_{t-i}=0)} + \sum_{i=1}^{q} \beta_i \, \ln \mu_{t-i} + \sum_{i=1}^{n^*} \varrho_i \, \pi_{t-i}.$$

In the resulting model, the intercepts ϖ and ω are not identified without additional restrictions, such as $\varpi = 0$. Alternatively, or additionally, dynamic spillover effects might be also modeled by the inclusion of the lagged endogenous variables of the two equations, see, e.g., Russell and Engle (2005) in an ACD-ACM context.

1.3.4 Empirical Evidence on DZA-MEM Processes

We apply a DZA-MEM by parameterizing the conditional mean function μ_t based on the Log-MEM specification (1.17). The lag orders in both dynamic components are chosen according to the Schwarz information criterion. Table 1.7 shows the estimation results for the DZA-MEM with autologistic binary choice component. For all stocks, the large volume indicator Δ_t has a positive impact on the subsequent trading probability, but only for IBM this effect is significant at a 5% level. However, the lagged trade indicators are significantly positive in almost every case. Thus, trade occurrences are positively autocorrelated, which is in line with empirical market microstructure research (see, e.g., Engle, 2000).

For every stock, all Q-statistics of the autologistic residuals

$$u_t := \frac{\mathcal{I}_t - \hat{\pi}_t}{\sqrt{\hat{\pi}_t \ (1 - \hat{\pi}_t)}}, \tag{1.34}$$

are significant at the 5% level, showing that an autologistic specification does not completely capture the dynamics and is too parsimonious.

As shown by Table 1.8, dynamic modeling of trade occurrences by an ACM specification yields significantly lower Q-statistics. Hence, the ACM specification seems to fully capture the serial dependence in the trade indicator series with the parameter estimates underlining the strong persistence in the process. For MDR, the smallest root of the polynomial $1 - \zeta_1 z - \zeta_2 z^2 = 0$ is not far outside the unit circle, while in the other cases, the coefficient ζ_1 is close to one, suggesting that the underlying process is very persistent.

1.3.5 Evaluating the DZA-MEM: Density Forecasts

The evaluation of the DZA-MEM is complicated by the fact that the disturbances are not i.i.d. In particular, the non-identical distribution makes an application of the semiparametric specification test from Section 1.2.3 impossible. Moreover, since the disturbances are not i.i.d. even given the restricted information set $\mathcal{H}_{t-1}$, we cannot employ a transformation that provides standardized i.i.d. innovations as in De Luca and Zuccolotto (2006).

As an alternative, we examine one-step-ahead forecasts of the conditional density of y_t implied by the DZA-MEM, which we denote by $f_{t|t-1}(y_t|\mathcal{F}_{t-1})$. To assess the forecasting performance of our model, we employ evaluation methods as developed by Diebold et al. (1998) and firstly applied to MEM-type models by Bauwens et al. (2004). One difficulty is that these methods are designed for continuous random variables, while we have to deal with a discrete probability mass at zero. Therefore, following Liesenfeld

Table 1.7: Estimation Results – DZA-MEM with Autologistic Component
Maximum likelihood estimates of the DZA-MEM based on the ZAF distribution with autologistic specification for the binary choice component. The Q-statistics are based on the residuals of the autologistic component. 5% (1%) critical values of the Q-statistics with 20, 50 and 100 lags are 31.41 (37.57), 67.51 (76.15) and 124.34 (135.81), respectively. The autologistic residuals are defined as: $u_t := \frac{\mathcal{I}_t - \hat{\pi}_t}{\sqrt{\hat{\pi}_t\,(1-\hat{\pi}_t)}}$.

	BAC		**IBM**		**MDR**		**XEC**	
	Coef.	T-St.	Coef.	T-St.	Coef.	T-St.	Coef.	T-St.
ω	0.047	6.522	0.036	8.104	0.044	6.067	0.005	1.577
α_1	0.120	11.135	0.206	9.691	0.182	14.880	0.153	8.669
α_2	-0.060	-4.407	-0.128	-5.685	-	-	-0.056	-3.039
β_1	0.908	58.869	0.919	115.177	0.854	70.964	0.923	118.600
α_1^0	-0.416	-6.146	-0.310	-3.895	-0.133	-8.873	-0.260	-7.412
α_2^0	0.345	4.939	0.223	2.981	-	-	0.218	6.058
m	1.755	10.270	653.760	4.193	450.064	11.143	507.708	9.435
η	562.562	8.006	7.719	22.007	5.393	13.260	2.729	14.331
a	0.560	17.809	0.378	52.429	0.493	24.486	0.862	25.311
θ_0	-0.390	-1.748	-0.937	-4.780	-1.440	-30.669	-1.196	-22.787
θ_1	0.080	1.788	0.087	2.987	0.009	1.430	0.021	1.550
γ_1	0.697	5.971	0.456	5.755	0.453	10.860	0.525	10.599
γ_2	0.591	5.313	0.217	3.743	0.359	8.981	0.213	5.874
γ_3	0.400	3.473	0.349	6.553	0.257	6.357	0.211	5.909
γ_4	0.719	6.755	0.299	5.584	0.305	7.717	0.164	4.595
γ_5	0.637	5.860	0.153	2.229	0.263	6.715	0.120	3.501
γ_6	-	-	0.115	1.227	0.154	3.942	0.209	5.959
γ_7	-	-	0.194	3.424	0.228	5.864	0.126	3.526
γ_8	-	-	0.177	2.446	0.259	6.599	0.209	5.875
γ_9	-	-	0.201	2.871	0.119	3.077	0.153	4.316
γ_{10}	-	-	0.164	1.912	0.203	5.184	0.125	3.596
γ_{11}	-	-	0.202	3.744	-	-	0.133	3.652
γ_{12}	-	-	0.204	4.038	-	-	0.210	6.009
$\mathcal{L}$	-9217.064		-10581.962		-10083.494		-10585.878	
SIC	18577.504		21370.032		20337.240		21377.865	
$Q(20)$	183.111		51.498		51.409		64.732	
$Q(50)$	446.024		247.462		167.692		227.053	
$Q(100)$	827.023		538.224		325.533		445.320	

Table 1.8: Estimation Results – DZA-MEM with ACM Component
Maximum likelihood estimates of the DZA-MEM based on the ZAF distribution with ACM specification for the binary choice component. The Q-statistics are based on the residuals of the ACM component. 5% (1%) critical values of the Q-statistics with 20, 50 and 100 lags are 31.41 (37.57), 67.51 (76.15) and 124.34 (135.81), respectively. The ACM residuals are defined as: $u_t := \frac{\mathcal{I}_t - \hat{\pi}_t}{\sqrt{\hat{\pi}_t\,(1-\hat{\pi}_t)}}$.

	BAC		**IBM**		**MDR**		**XEC**	
	Coef.	T-St.	Coef.	T-St.	Coef.	T-St.	Coef.	T-St.
ω	0.047	7.378	0.034	10.855	0.037	8.048	0.012	5.676
α_1	0.117	10.941	0.186	12.419	0.107	10.055	0.130	7.914
α_2	-0.060	-4.478	-0.122	-7.589	-	-	-0.063	-3.629
β_1	0.913	68.542	0.935	157.746	0.925	97.593	0.950	210.962
α_1^0	-0.409	-6.243	-0.290	-7.392	-0.092	-9.030	-0.215	-6.563
α_2^0	0.311	4.577	0.199	5.054	-	-	0.170	5.128
m	1.701	11.707	653.999	3.791	452.493	5.375	507.657	11.273
η	562.562	9.848	7.523	10.040	3.636	9.763	2.249	11.319
a	0.570	20.125	0.385	19.177	0.618	18.001	0.972	19.707
ϖ	0.018	3.304	0.006	2.946	0.000	-0.855	0.001	0.574
ρ_1	0.195	5.835	0.183	7.911	0.146	10.818	0.203	9.696
ρ_2	-0.077	-2.267	-0.099	-4.325	-0.132	-9.907	-0.125	-5.945
ζ_1	0.993	501.627	0.995	664.373	1.806	116.644	0.993	683.977
ζ_2	-	-	-	-	-0.807	-51.669	-	-
$\mathcal{L}$	-9114.437		-10475.223		-9969.811		-10453.388	
SIC	18345.367		21066.941		20047.150		21023.271	
$Q(20)$	36.885		34.602		37.773		16.338	
$Q(50)$	71.301		55.415		75.476		33.972	
$Q(100)$	128.234		104.017		114.527		97.845	

et al. (2006) and Brockwell (2007), we employ a modified version of the test. The idea is to add random noise to the discrete component, ensuring that the c.d.f. is invertible. Hence, we compute randomized probability integral transforms (PITs)

$$z_t := \begin{cases} U_t \; F_{t|t-1}(y_t|\mathcal{F}_{t-1}) & \text{if} \quad y_t = 0, \\ F_{t|t-1}(y_t|\mathcal{F}_{t-1}) & \text{if} \quad y_t > 0, \end{cases} \tag{1.35}$$

where $F_{t|t-1}(y_t|\mathcal{F}_{t-1})$ denotes the c.d.f. corresponding to $f_{t|t-1}(y_t|\mathcal{F}_{t-1})$, while U_t are random variables with $\{U_t\}_{t=1}^n$ being i.i.d. $U(0,1)$. Using (1.22), we obtain

$$z_t = \begin{cases} U_t \; (1-\pi_t) & \text{if} \quad y_t = 0, \\ (1-\pi_t) + \pi_t \, G_{t|t-1}(y_t/\mu_t|\mathcal{H}_{t-1}) & \text{if} \quad y_t > 0, \end{cases} \tag{1.36}$$

where $G_{t|t-1}(y_t/\mu_t|\mathcal{H}_{t-1})$ is the c.d.f. corresponding to $g_{t|t-1}(y_t/\mu_t|\mathcal{H}_{t-1})$, which denotes the one-step-ahead forecast of the conditional density of the disturbance ε_t for $\varepsilon_t > 0$ evaluated at y_t/μ_t. For a DZA-MEM based on the ZAF distribution, it follows that

$$z_t = \begin{cases} U_t \; (1-\pi_t) & \text{if} \quad y_t = 0, \\ (1-\pi_t) + \pi_t \; [\mathcal{B}(c\,; m, \eta) \,/ \mathcal{B}(m, \eta)] & \text{if} \quad y_t > 0, \end{cases} \tag{1.37}$$

where $\mathcal{B}(c\,; m, \eta) := \int_0^c t^{m-1} \, (1-t)^{\eta-1} \, dt$ is the incomplete beta function evaluated at

$$c := \left(y_t \, \mu_t^{-1} \, \pi_t \, \xi\right)^a \left[\eta + \left(y_t \, \mu_t^{-1} \, \pi_t \, \xi\right)^a\right]^{-1}. \tag{1.38}$$

If the series of one-step-ahead forecasts, $f_{t|t-1}(y_t|\mathcal{F}_{t-1})$, coincides with the true conditional densities, $f_Y(y_t|\mathcal{F}_{t-1})$, the z_t sequence is i.i.d. $U(0,1)$, see Brockwell (2007) for a proof. While Diebold et al. (1998) recommend a visual inspection of the properties of the z_t's, we also check for uniformity using the Pearson-χ^2 and Kolmogorov-Smirnov (KS) tests. In addition, following Berkowitz (2001), we compute the normal quantile transformation $z_t^{\text{tr}} := \Phi^{-1}(z_t)$, where $\Phi^{-1}(\cdot)$ denotes the inverse c.d.f. of the standard normal distribution. As is well-known, i.i.d. uniformity of the z_t's implies that the z_t^{tr} sequence is i.i.d. $N(0,1)$. To verify normality, we consider the omnibus tests proposed by Bowman and Shenton (1975) and Doornik and Hansen (2008), which will be referred to as BS and DH test, respectively.

Appendix A.1 describes the setup and reports the results of a power study for the above distribution tests based on both in-sample and out-of-sample PITs. In the latter case, estimation is carried out using the first two thirds of the dataset, while density forecasts and PITs are computed for the last third of the sample. The results can be summarized as follows. First, the power with respect to misspecifications of the error distribution is high for all tests with rejection rates being close or equal to one. Second, the detection of misspecifications of zero dynamics is somewhat lower. However, the power of all tests increases substantially when evaluating out-of-sample instead of in-sample density forecasts. Finally, there are noticeable performance differences between the four tests. In the in-sample setting, the BS and DH tests offer the highest power, while the KS test performs relatively poorly. For out-of-sample forecasts, the KS test

Table 1.9: Distribution Tests for (Transformed) Out-of-Sample PITs
P-values of the χ^2 test and statistic of the Kolmogorov-Smirnov test (KS) for uniformity of out-of-sample randomized PITs. 5% and 1% critical values of the KS test are 0.027 and 0.032, respectively. In addition, p-values of the Bowman-Shenton (BS) and Doornik-Hansen test (DH) for normality of transformed randomized PITs are reported. The estimated model is the DZA-ACM-MEM. Two thirds of the sample are used for estimation, one third for evaluation.

	BAC	**IBM**	**MDR**	**XEC**
		Uniformity of PITs		
$\chi^2_{\text{P-Val}}$	0.243	0.818	0.887	0.752
KS_{Stat}	0.022	0.013	0.024	0.017
		Normality of Transformed PITs		
$\text{BS}_{\text{P-Val}}$	0.100	0.474	0.713	0.385
$\text{DH}_{\text{P-Val}}$	0.109	0.463	0.765	0.414

becomes the most powerful one. Due to the reported power gains in this setting, but also motivated by the higher practical relevance, we focus on the evaluation of out-of-sample forecasts in the empirical application below.[2]

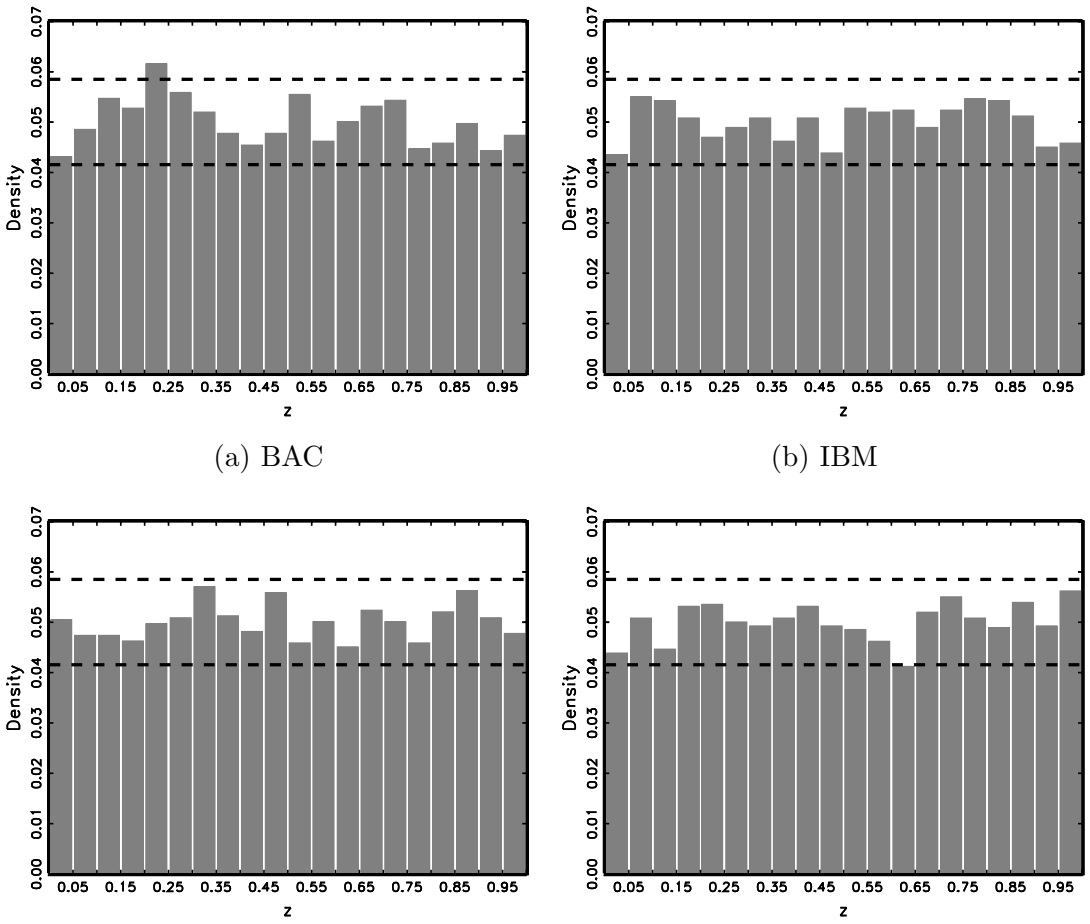
Table 1.9 shows the results of the distribution tests for the out-of-sample randomized PITs and their transformed counterparts implied by the DZA-ACM-MEM. As above, the model was estimated using the first two thirds of the sample, while density forecasts were computed for the last third. For all stocks and both the χ^2 and KS test, we cannot reject the null hypothesis of uniformity of the PITs at a significance level of 5%. Similarly, the BS and DH statistic, checking for normality of the transformed PITs, are insignificant at the 5% level in all cases. These findings are underlined by the histograms of the out-of-sample PITs depicted in Figure 1.6. For all stocks, most bars are well within the 95% confidence bounds, which indicates a satisfactory density forecasting performance.

1.4 Conclusion

We propose a model for autoregressive positive-valued variables with excess zero outcomes. These properties are typical for time-aggregated financial high-frequency data and cannot be appropriately handled in extant approaches.

In order to capture observations clustered at zero, we introduce a new point-mass mixture distribution, which consists of a discrete component at zero and a flexible con-

[2]Results of in-sample forecasts can be found in a web appendix available at `http://amor.cms.hu-berlin.de/~malecpet/ZAMEM_appendix.pdf`.



(a) BAC (b) IBM

(c) MDR (d) XEC

Figure 1.6: Histograms of Out-of-Sample PIT Sequences
Histograms of the out-of-sample randomized probability integral transforms based on the estimated DZA-ACM-MEM. Two thirds of the entire sample are used for estimation, one third for evaluation. The dashed lines represent approximate 95% confidence intervals for the bin heights under the null that the PITs are i.i.d. uniform.

tinuous distribution for the strictly positive part of the support. To evaluate such a distribution, a novel semiparametric specification test tailored for point-mass mixture distributions is introduced. Finally, to accommodate serial dependencies in the data, we incorporate the proposed point-mass mixture into a new type of multiplicative error model (MEM) capturing the dynamics of both zero occurrences and strictly positive values. In a simulation study, we demonstrate that, in the presence of zero observations, maximum likelihood estimation of the resulting zero-augmented MEM (ZA-MEM) offers clear efficiency gains compared to exponential QML and that the proposed specification test exhibits excellent power in detecting misspecifications of the error distribution.

Empirical evidence based on cumulated trading volumes of four NYSE stocks shows that the zero-augmented MEM on the basis of the proposed point-mass mixture captures the distributional and dynamic properties of the data very well. The best fit is shown for a specification incorporating a two-state ACM component for the trade indicator. Besides MEM dynamics in trading volumes, the model also explains individual dynamics in trade occurrences and produces good out-of-sample density forecasts.

Further possible applications include the modeling of absolute returns revealing a nontrivial proportion of zero outcomes or the modeling of irregularly-spaced high-frequency data, where zero durations occur as a consequence of simultaneous transactions. An alternative motivation for continuous-discrete mixture distributions is, for instance, the clustering of trade sizes at round numbers, which is caused by the well-known preference of traders for round lot sizes.

Finally, our modeling framework is sufficiently flexible to be extended in various ways, e.g., to allow for dynamic spillovers between the two types of dynamics or incorporating other exogenous regressors. Moreover, it should be straightforward to extend the model to a multivariate setting, in the spirit of, e.g., Manganelli (2005), Cipollini et al. (2006) or Hautsch (2008). Here, the modeling of equidistant high-frequency data is particularly useful, as the regular sampling grid avoids the technical complications caused by the asynchronicity of observations. An interesting application of a multivariate extension would be the simultaneous analysis of trading volumes on several exchanges. In this context, the occurrence of zero volumes on specific venues only could be interpreted as a substitution effect, while zero volumes on all exchanges would indicate a complete lack of information.

Chapter 2

Nonparametric Kernel Density Estimation Near the Boundary

This chapter is based on Malec and Schienle (2012).

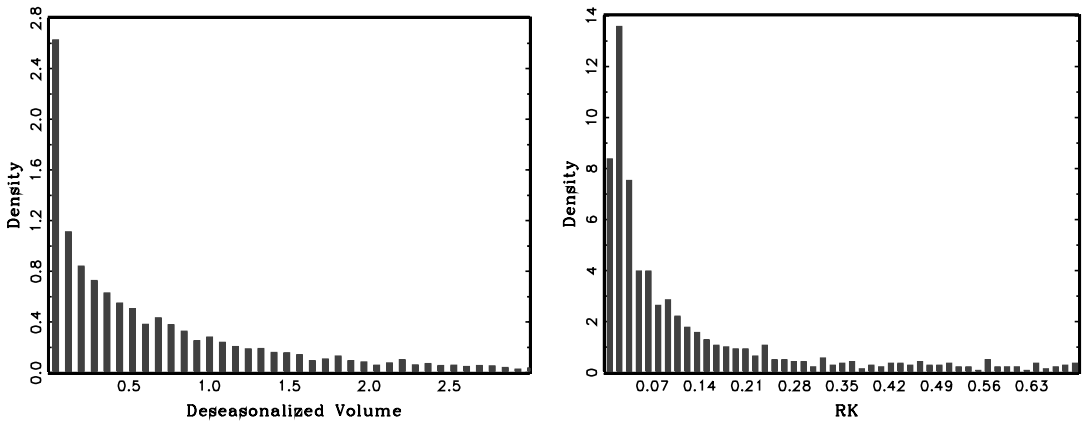
2.1 Introduction

There are many applications, in particular in economics, where densities of positive-valued random variables are the object of interest or an essential model ingredient to be estimated from a given sample. Compare, e.g., data on incomes, survival times, financial trading volumes and durations, as well as volatility measures. In many of these situations, however, appropriate functional forms are unknown or controversial, such that a nonparametric estimate is needed. Importantly, it is often the point estimates close to the boundary which are in the focus of practical interest and thus require good precision.

In case of densities where most of the probability mass is concentrated away from the boundary, there is a huge literature on boundary correction techniques for the standard symmetric fixed kernel density estimator. Such adjustments are needed at points close to the boundary, since fixed kernels might assign positive weight outside the support, yielding inconsistent results. Among these techniques count, e.g., the cut- and normalized kernel, see Gasser and Müller (1979), and the reflection method, see Schuster (1958).

If, however, the true density might have substantial mass close to the boundary, there are superior methods, such as the boundary kernel of Jones (1993). Since this estimator could yield negative point estimates, Jones and Foster (1996) propose an appropriate correction at some minor cost of performance (see Jones, 1993). In comparison, the combination of polynomial transformation followed by reflection as in Marron and Ruppert (1994) is much less flexible, working well at boundaries only if the initial transformation is close enough to the density shape near zero.

Nonparametric kernel density estimators based on asymmetric kernels, such as those of gamma-type, have been introduced to improve upon the performance of fixed kernels



(a) Volume (b) Realized Kernel

Figure 2.1: Histograms of Intraday Trading Volume and Realized Kernel Estimates
We consider deseasonalized nonzero 15-second trading volumes of Citigroup and realized kernel estimates for JP Morgan. Sample period: February 2009 (trading volumes), January 2006 – December 2009 (realized kernel). For details on the seasonal adjustment of trading volumes and the computation of the realized kernel, see Section 2.4.

at the boundary. In particular for positive random variables, their flexible shape avoids the boundary consistency problem and directly yields positive estimates by construction (see Chen, 2000). Moreover, in this class of nonnegative kernel density estimators, asymmetric kernels achieve the optimal rate of convergence in the sense of the integrated mean squared error (IMSE) (see, e.g., Chen, 2000; Scaillet, 2004). In addition, their variance decreases the further the points of estimation move away from the boundary. This effect leads to an advantage in situations of naturally unbalanced scattered design points, in particular for densities with sparse areas (see, e.g., Chen, 1999; Hagmann and Scaillet, 2007; Michels, 1992). As generally boundary and unequal design issues get increasingly severe for higher dimensions, the use of gamma kernels especially pays off for multivariate density or regression problems. This is relevant, in particular, for the extreme case of functional data analysis (see Ferraty and Vieu, 2006).

We contribute to the extensive literature on kernel estimation near the boundary by clearly identifying design situations in which the finite sample and asymptotic performance of gamma kernel estimators is distinctly superior to any competing fixed kernel adjustments, implying that the former should be strictly preferred. Such situations occur when the true density f approaches the boundary with a derivative f' significantly different from zero. These density shapes naturally appear in high-frequency data, e.g., when studying cumulated trading volumes (see Figure 2.1), but also in many other applications, such as spectral density estimation of long-memory time series (see, e.g., Robinson and Henry, 2003) or when modeling volatilities, in particular at the intraday level (see, e.g., Corradi et al., 2009).

We also show that, depending on the underlying shape of the true density, the two existing gamma kernel estimators, the so-called standard and modified version introduced in Chen (2000), might differ substantially in terms of boundary performance and still leave significant room for improvement. While in practice almost exclusively a modified gamma-type kernel estimator is used, we find that, in particular for pole situations, the standard gamma-type estimator yields large performance advantages. We therefore introduce a simple data-driven criterion identifying such extreme settings.

For all other design situations, we propose a refined gamma kernel estimator, which outperforms all existing estimators in a comprehensive finite sample study. The new estimator introduces a modification parameter according to the shape of f and its first two derivatives close to the boundary. For determining the appropriate specification of this refined gamma kernel estimator in practice, we additionally provide an automatic procedure.

Our two applications clearly demonstrate the significant impact of a design-dependent choice of the gamma kernel type on the overall estimation results. For high-frequency stock trading volumes, we detect a pole situation and obtain an improved fit employing the standard gamma kernel estimator, as opposed to the generally applied modified one. In realized volatility modeling, the new refined gamma kernel estimator is the only one which yields results consistent with financial theory, while all other competing estimators produce an unexpected bias.

2.2 Kernel Density Estimation at the Boundary

Throughout the chapter, we study density estimation for the case that the support $S_X \subset \mathbb{R}$ of an unknown density is bounded from one side. Without loss of generality, we assume that this bound is a lower bound and equals zero, such as in applications involving the distributions of wages, trading volumes, etc. Obtained results, however, can be easily generalized by appropriate translations and reflections at the y-axis. In addition, note that we restrict our exposition to the case of univariate densities for ease of notation. Multivariate extensions are systematically straightforward via product kernels.

For an i.i.d. random sample $\{X_i\}_{i=1}^n$ from a distribution with unknown density $f_X(x)$, the conventional kernel density estimator has the form

$$\hat{f}_X(x) = \frac{1}{nb} \sum_{i=1}^{n} K\left(\frac{x - X_i}{b}\right), \tag{2.1}$$

where b denotes a smoothing bandwidth with $b \to 0$ and $nb \to \infty$ as $n \to \infty$, while K is a kernel function which integrates to unity, i.e., $\int K(u)\,du = 1$. If the shape of K is symmetric and fixed across the support, estimation and inference are generally simplified given an unbounded support. However, if zero bounds the support S_X from below, $\hat{f}_X$ is inconsistent in the boundary region $[0, b)$ for such simplistic choices of K. Therefore, the literature has provided many suggestions for adjustments in fixed kernel estimation,

which we will outline in more detail when they appear as benchmarks in the simulation study in Section 2.3. What characterizes all these approaches, however, is that, predominantly, they work well only for specific forms of f_X in the boundary region and/or can yield negative estimates. In particular, for densities with nonvanishing probability mass close to zero as in Figure 2.1, these standard correction methods perform poorly at the boundary. In applications, it is exactly the boundary region which lies in the focus of attention and requires precise estimates, though.

2.2.1 Standard Asymmetric Kernel Density Estimators

Density estimators based on kernels with locally varying form have shown good performance for a wide range of shapes of the underlying true density. Such kernels are nonnegative, but no longer symmetric, adjusting in skewness along the support. For the considered one-sided boundary problem, gamma kernel estimators are the simplest and most popular forms of such flexible estimators. In case of a two-sided boundary, which is not our focus here, beta kernels would be the appropriate choice (see Chen, 1999). There are two alternative specifications of gamma kernel estimators proposed by Chen (2000), of which the first kind is defined as

$$\hat{f}_X^{\gamma}(x) = \frac{1}{n} \sum_{i=1}^{n} K_{x/b+1,b}^{\gamma}(X_i)\,, \tag{2.2}$$

where $K_{x/b+1,b}^{\gamma}$ denotes the density of the gamma distribution with shape parameter $x/b+1$ and scale parameter b, i.e.,

$$K_{x/b+1,b}^{\gamma}(u) := \frac{u^{x/b} \exp(-u/b)}{b^{x/b+1}\,\Gamma(x/b+1)}. \tag{2.3}$$

Consistency and asymptotic normality of the above estimator are straightforward to derive under standard assumptions. See, e.g., Chen (2000) for the pointwise and Hagmann and Scaillet (2007) for the uniform version. For time series observations, consistency can also be obtained under mixing assumptions following Bouezmarni and Rombouts (2010). In particular, for a sufficiently smooth density $f_X \in C^2(S_X)$, it can be shown that bias and variance vanish asymptotically for $b \to 0$ and $nb \to \infty$. Their asymptotic forms are

$$\mathrm{Bias}\Big\{\hat{f}_X^{\gamma}(x)\Big\} = b\,\left\{f_X'(x) + \frac{1}{2}\,x\,f_X''(x)\right\} + o(b)\,, \tag{2.4}$$

and

$$\mathrm{Var}\Big\{\hat{f}_X^{\gamma}(x)\Big\} \approx \begin{cases} \frac{f_X(x)}{nb}\,\mathcal{C}_b(x) & \text{if} \quad x/b \to \kappa, \\ \frac{f_X(x)}{2\sqrt{\pi}}\,(xb)^{-1/2}\,n^{-1} & \text{if} \quad x/b \to \infty, \end{cases} \tag{2.5}$$

where κ is a nonnegative constant and

$$\mathcal{C}_b(x) := \frac{\Gamma(2\kappa+1)}{2^{1+2\kappa}\,\Gamma^2(\kappa+1)}. \tag{2.6}$$

Accordingly, the asymptotic mean squared error is

$$\mathrm{MSE}\Big\{\hat{f}_X^{\gamma}(x)\Big\} \approx \begin{cases} b^2\,\left\{f_X'(x)+\frac{1}{2}\,x\,f_X''(x)\right\}^2 + \frac{f_X(x)}{nb}\,\mathcal{C}_b(x) & \text{if} \quad x/b \to \kappa, \\ b^2\,\left\{f_X'(x)+\frac{1}{2}\,x\,f_X''(x)\right\}^2 + \frac{f_X(x)}{2\sqrt{\pi}}\,(xb)^{-1/2}\,n^{-1} & \text{if} \quad x/b \to \infty. \end{cases} \tag{2.7}$$

Note that the asymptotic variance decreases for large x, which is offset by an increasing bias. In contrast to fixed kernel estimators, the asymptotic bias contains the first derivative of the density, f_X', which is due to the fact that the chosen flexible kernel shape has its mode rather than its mean at the point of estimation x. The modified gamma kernel estimator improves on this for most of the support without generating convergence problems in the boundary region. In particular, it uses the density of a gamma distribution with shape parameter x/b and scale parameter b as kernel function in the interior of the support. This density has mean x, but is unbounded for x approaching zero. Therefore, the kernel function consists of two regimes with the boundary form being chosen ad hoc to ensure a smooth connection to the desired interior shape, while avoiding unboundedness problems. According to Chen (2000), the estimator is thus defined as

$$\hat{f}_X^{\gamma m}(x) = \frac{1}{n}\sum_{i=1}^{n} K^{\gamma}_{\rho_b(x),b}(X_i)\,, \tag{2.8}$$

where

$$\rho_b(x) := \begin{cases} \frac{1}{4}\left(\frac{x}{b}\right)^2 + 1 & \text{if} \quad x \in [0,2b)\,, \\ x/b & \text{if} \quad x \in [2b,\infty)\,. \end{cases} \tag{2.9}$$

Note that the estimator fixes the size of the boundary region to the area from 0 to $2b$ independent of the shape of the underlying true density. The asymptotic bias of the modified gamma kernel estimator has the desired leading term, i.e.,

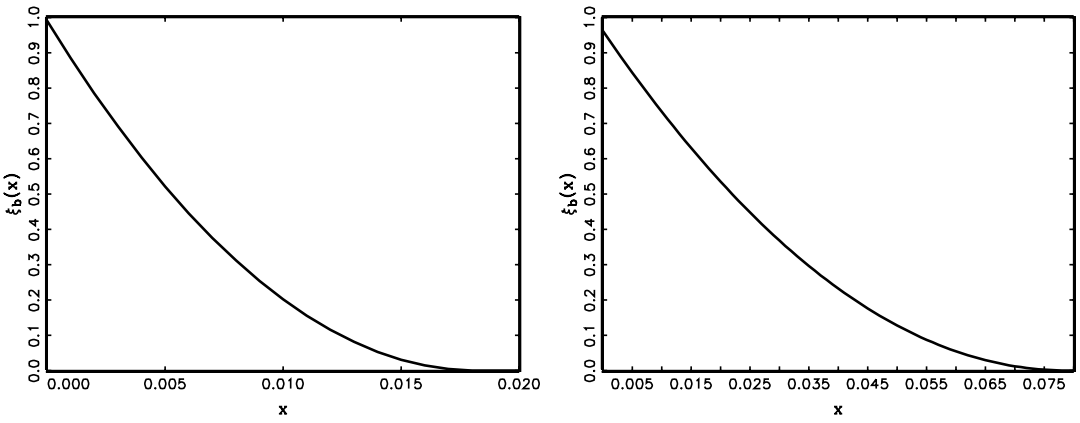
$$\mathrm{Bias}\Big\{\hat{f}_X^{\gamma m}(x)\Big\} = \begin{cases} \xi_b(x)\,b\,f_X'(x) + o(b) & \text{if} \quad x \in [0,2b)\,, \\ \frac{1}{2}\,x\,f_X''(x)\,b + o(b) & \text{if} \quad x \in [2b,\infty)\,, \end{cases} \tag{2.10}$$

where

$$\xi_b(x) := (1-x)\,\frac{\rho_b(x) - x/b}{1 + b\,\rho_b(x) - x}, \tag{2.11}$$

which is in $[0,1]$ for standard choices of $b < 1/2$ for all $x \in [0,2b)$ (see Figure 2.2). The corresponding asymptotic variance can be shown to have the same structure as (2.5) with modified constant

$$\widetilde{\mathcal{C}}_b(x) := \frac{\Gamma\big(2\kappa^2+1\big)}{2^{1+2\kappa^2}\,\Gamma^2(\kappa^2+1)}, \tag{2.12}$$



(a) $b = 0.0091$ (b) $b = 0.0396$

Figure 2.2: $\xi_b(x)$
Scale factor $\xi_b(x)$, as defined in (2.11), entering asymptotic bias of the modified gamma kernel estimator. Bandwidths of two DGPs from the simulation study in Section 2.3 are used.

and

$$\mathrm{MSE}\Big\{\hat{f}_X^{\gamma m}(x)\Big\} \approx \begin{cases} \{\xi_b(x)\, b\, f_X'(x)\}^2 + \frac{f_X(x)}{nb}\, \widetilde{\mathcal{C}}_b(x) & \text{if} \quad x/b \to \kappa, \\ \big\{\frac{1}{2}\, x\, f_X''(x)\, b\big\}^2 + \frac{f_X(x)}{2\sqrt{\pi}}\, (xb)^{-1/2}\, n^{-1} & \text{if} \quad x/b \to \infty. \end{cases} \tag{2.13}$$

See Chen (2000) for details on the derivations.

2.2.2 Choice of Estimators for Different Density Shapes Near Zero

Generally in the literature, the modified gamma kernel estimator has been strictly preferred to the standard gamma kernel version. Although a simple comparison of their asymptotic variances reveals that, close to the boundary (and for all $\kappa < 1$), the constant (2.12) for the modified estimator is strictly larger than the corresponding one for the standard gamma kernel in (2.6), the above choice has been justified by the similarity of the modified gamma kernel to fixed kernels in terms of asymptotic bias behavior as displayed in (2.10). However, when carefully comparing the leading asymptotic bias terms of both gamma-type estimators, we find that there are cases where it is asymptotically favorable to use the standard gamma kernel. For all $x > 2b$, in the interior of the support with

$$\left|\frac{1}{2}\, x\, f_X''(x)\right| > \left|f_X'(x) + \frac{1}{2}\, x\, f_X''(x)\right|, \tag{2.14}$$

the standard gamma kernel is preferable to the modified version. In particular, this occurs for areas where the density satisfies the shape restriction

$$0 < -f'_X(x)\,/\,f''_X(x) < x\ . \tag{2.15}$$

The lower bound is fulfilled for values x where f'_X and f''_X have different sign, i.e., where the density f_X is either decreasing and convex or where it is concave and increasing. In the first case, it can be shown that if f_X has a pole at zero, then trivially also the upper bound in (2.15) is satisfied. If additionally f_X does not have any local maxima, the standard gamma kernel should be preferred over the modified version for the entire interior part of the support (see Figure 2.3). Our simulation study below confirms that this effect is of importance in finite samples, as well, while being particularly pronounced when smaller sample sizes are considered. It can be easily shown that a pole is a sufficient condition, but the same logic also applies to all densities with $f'_X < -\alpha < 0$ for α not too small and $f''_X \geq 0$ close to the boundary.

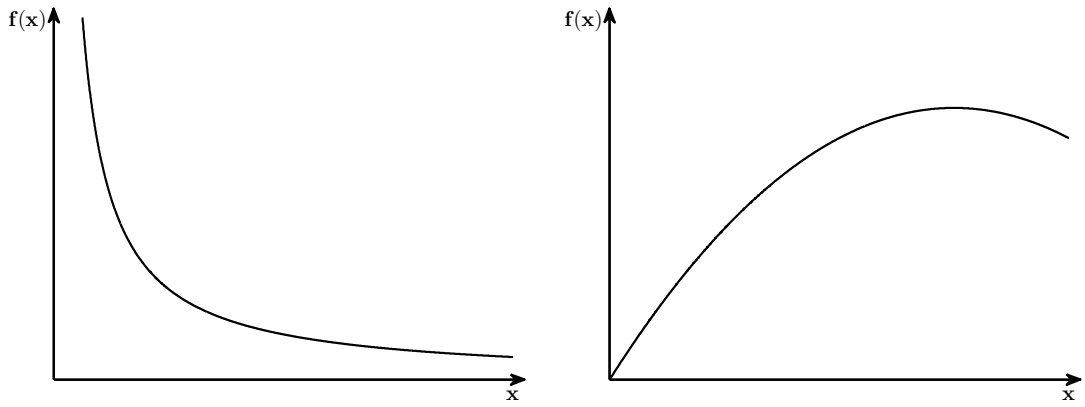
Apart from these pronounced cases at the boundary, any density whose support is unbounded from the right will be convex and decreasing for large x in order to be integrable. In this situation, the asymptotic variance regimes are identical for both gamma-type estimators. For the asymptotic bias, independent from the rate of decay of f_X, the upper bound in (2.15) always holds in these regions. For very large x, however, slopes and curvature values are generally small, yielding overall small biases for any kernel-type estimator, such that a measurable advantage of the standard versus modified gamma kernel estimator might disappear. Besides these convex cases, unimodal densities are concave around the mode and increasing to its left (see Figure 2.3). The use of the standard gamma kernel estimator might be recommendable in this area, as well. For finite samples, however, observed differences are rather small even in the extreme case of a strictly concave density between zero and the mode.

Moreover, also on the boundary for $x \in [0, 2b)$, the standard gamma kernel estimator can outperform the modified one if

$$\left|f'_X(x)\ \xi_b(x)\right| > \left|f'_X(x) + \frac{1}{2}\,x\,f''_X(x)\right|. \tag{2.16}$$

As $\xi_b(x) < 1$, this can occur for densities f_X with opposite sign of f'_X and f''_X. Thus in some pole situations satisfying (2.15), the standard gamma kernel is superior to the modified one due to $|\xi_b(x)|\,/\,\left|1 + \frac{1}{2}\,x\,f''_X(x)\,/f'_X(x)\right| > 1$ with $\xi_b(x) = (1-\beta\,b)\,/\,(1+b+c_\beta)$, where $c_\beta > 0$ for any $x = \beta\,b$ in the boundary region. In particular, the above condition is fulfilled for densities f_X which can be approximated by $\delta\,x^{-1}$ or $\delta\,x^{-2}$ for $\delta > 0$ in the boundary region. However, as this area is vanishingly small, its influence on the overall estimation results is negligible (c.f. the simulation results in Section 2.3).

Hence in practice, the ex-ante detection of pole situations is crucial for being able to choose the best-performing version among standard and modified gamma kernel estimators. We propose a simple but reliable measure to check for poles as opposed to standard cases. If f_X has a pole at zero, it is the relative convergence and consistency



(a) Pole at $x = 0$ and no local maxima (b) Unimodal

Figure 2.3: Density Shapes Favoring the Standard Gamma Kernel Estimator
Schematic densities for which the standard gamma kernel estimator in (2.2) and (2.3) should be preferred over the modified version in (2.8) and (2.9) according to the shape restrictions (2.15) and (2.16). Left figure: condition (2.15) *is satisfied globally* for $x > 2b$, while for $x \leq 2b$, condition (2.16) holds if f_X can be bounded by $\delta\, x^{-3}$ with $\delta > 0$ in this area. Right figure: condition (2.15) *can be satisfied locally* to the left of the mode.

of the estimator $\hat{f}_X$ which is of main importance to judge if the correct order of decay is detected.[1] Thus, it must hold that

$$\left| \hat{f}_X(x) \,/ f_X(x) - 1 \right| = o_p(1)\,. \tag{2.17}$$

The governing term in the stochastic expansion for the right-hand side controlling convergence is $x\, f'_X(x) \,/ f_X(x)$, which we write as $x\, D(x)$. See the proof of Theorem 5.3. in Bouezmarni and Scaillet (2005). The practically most important pole situations occur for densities which have or can be bounded by densities with hypergeometric decay from zero, i.e., $f_X(x) = \zeta\, x^{-\vartheta}$ for $\zeta > 0$ and $0 < \vartheta < 1$ (the cases with $\vartheta > 1$ are excluded by f_X being a density). Here, the quantity $x\, D(x)$ equals the constant $-\vartheta$ irrespective of the scaling factor ζ.

For distinguishing a pole situation from a no-pole scenario, it is favorable to study $D(x)$ directly in order to ensure sufficient power of the criterion against alternatives. Therefore, we estimate $D(x)$ by exploiting the simple relation

$$D(x) := \frac{f'_X(x)}{f_X(x)} = \frac{d}{dx} \ln f_X(x). \tag{2.18}$$

Note that, for x approaching 0, in a pole situation $D(x)$ is significantly negative, approaching infinity at rate $-\vartheta/x$ in case of densities decreasing with hypergeometric speed

[1] See, e.g., Robinson and Henry (2003) for how this is important regarding consistent estimation of the long-memory parameter in long-range dependent time series.

and -1 for exponential-type behavior. In all other settings where the modified gamma kernel is the method of choice, $D(x)$ is significantly positive. As a criterion, $D(x)$ combines properties of the density and its slope to distinguish the pole situation from other density shapes. This is more powerful than checking density and slope in isolation. In practice, $D(x)$ can be estimated by the difference quotient based on modified gamma kernels, i.e.,

$$\widehat{D}(x) = \frac{\ln \hat{f}_X^{\gamma m}(x+b) - \ln \hat{f}_X^{\gamma m}(x)}{b}, \tag{2.19}$$

where $b > 0$ is the same bandwidth as for the density estimates at x and $x + b$. For the practical framework of this chapter, it is sufficient to work with a rough criterion checking if $\widehat{D}(x)$ is significantly negative or not. Developing a novel formal test for the H_0 of a hypergeometric pole situation is beyond the scope of this chapter. We conjecture that, using the results in Fernandes and Grammig (2005) for specification testing in the simple density case, the corresponding asymptotic distribution of the centered test statistic

$$T_D := nb^2 \left(\widehat{D}(x) + \frac{\vartheta}{x} \right), \tag{2.20}$$

could be derived. However, as calculations are quite involved and should be complemented with a valid bootstrap approximation scheme for finite samples, we leave this for future research and a study on its own.

2.2.3 Refined Estimation with Modified Gamma Kernels

In cases where we can exclude a pole at the boundary, the modified gamma kernel generally should be the method of choice in terms of best asymptotic performance. However in the literature, its chosen form, in particular in the boundary region, has mainly been justified by (computational) convenience. Our simulation results clearly indicate that alternative, slightly more flexible specifications can significantly improve upon the performance of the original modified gamma kernel.

More precisely, we propose simple refined versions of the modified gamma kernel, where an additional modification parameter c allows for higher accuracy if appropriately chosen in a data-driven way. We study two types of refined modified gamma kernels, i.e.,

$$\rho_b^{v_{\mathrm{I}}}(x) := \begin{cases} \left[\frac{1}{4} \left(\frac{x}{bc} \right)^2 + 1 \right] [c + 2b(1-c)] & \text{if} \quad x \in [0, 2bc), \\ \frac{x}{bc} (c + 2b - x) & \text{if} \quad x \in [2bc, 2b), \\ x/b & \text{if} \quad x \in [2b, \infty), \end{cases} \tag{2.21}$$

and

$$\rho_b^{v_{\mathrm{II}}}(x) := \begin{cases} \frac{1}{4} \left(\frac{x}{bc} \right)^2 + 1 & \text{if} \quad x \in [0, 2bc), \\ x/(bc) & \text{if} \quad x \in [2bc, \infty), \end{cases} \tag{2.22}$$

where $c \in (0,1]$ with $c = 1$ yielding the original parametrization in both cases. Specification v_{I} shifts the boundary regime below one and introduces a flexible quadratic middle part. In the latter regime, for $\rho_b(x) > x/b$, we have that $x/b < \rho_b^{v_{\text{I}}}(x) < \rho_b(x)$ if

$$\frac{x}{b}\,\frac{2b-x}{\rho_b(x)-x/b} < c < 1, \qquad x \in [2bc, 2b)\,, \tag{2.23}$$

where $\rho_b(x)$ is defined as in (2.9). Importantly, fulfillment of the above condition implies that specification v_{I} is closer to the theoretically optimal situation with the mean of the kernel being at the observation point as compared to the original modified gamma kernel. The second alternative, v_{II}, keeps two regimes and the general structure of the original specification but shrinks the boundary region proportionally to the value of the modification parameter c. This modification also affects asymptotics in the interior of the support, as the mean of the kernel equals x/c and hence only in the trivial case $c = 1$ coincides with the point of estimation.

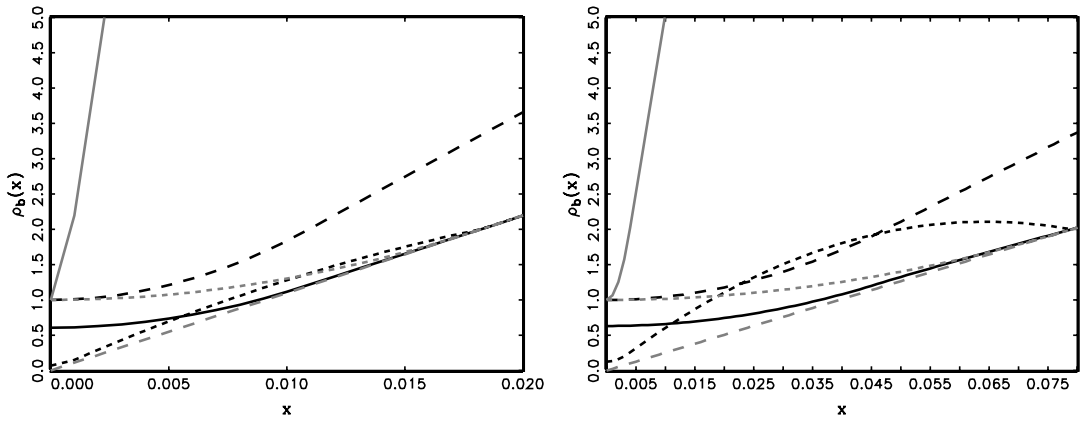
Figure 2.4 shows plots of $\rho_b(x)$ based on the original specification proposed by Chen (2000) along with the above refined versions for different values of the parameter c and using the bandwidths of two DGPs from the simulation study in Section 2.3. In addition, we include x/b, which corresponds to the interior component of the original specification and implies a gamma kernel with mean at the point of estimation. In its middle regime, $\rho_b^{v_{\text{I}}}$ is closer to x/b than the original specification for $c = 0.6$ in the right and for both values of c in the left figure, since in these cases, condition (2.23) is satisfied. Close to the boundary, the shape function of specification v_{I} takes values below one, implying that the resulting gamma densities and thus gamma kernels are unbounded at the origin (see Figure 2.5). However, the finite sample study below clearly reveals that this specification outperforms both the original modified one and the refined version v_{II} in all settings where a modified gamma kernel should be applied.

For a feasible implementation of these refined estimators, we provide an automatic procedure to select the modification parameter c. Holding the bandwidth b fixed, we determine the threshold $x^c := b\,\kappa$ for which the two MSE expressions of the modified gamma kernel in (2.13) coincide. Then, the optimal value of c can be obtained as $c^* = \kappa/2 = x^c/(2b)$. In practice, this approach requires minimizing the objective function

$$\begin{aligned}\mathcal{M}(x) := \Bigg\{ &\left[\xi_b(x)\; b\, f_X'(x)\right]^2 + \frac{f_X(x)}{nb}\,\widetilde{\mathcal{C}}_b(x) - \left[\frac{1}{2}\,x\, f_X''(x)\, b\right]^2 \\ &- \frac{f_X(x)}{2\sqrt{\pi}}\,(xb)^{-1/2}\,n^{-1}\Bigg\}^2\,,\end{aligned} \tag{2.24}$$

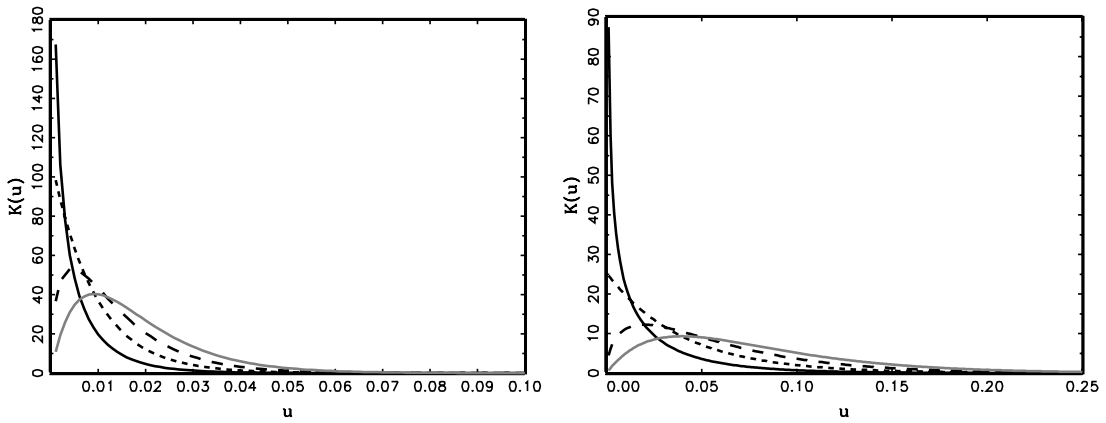
in $0 \le x \le 2b$.

Evaluation of the objective function requires estimates of the unknown density and its first two derivatives. $f_X(x)$ and $f_X'(x) = D(x)\; f_X(x)$ can be estimated using the original modified gamma kernel. An estimate of $f_X''(x)$ can be obtained by differentiating, e.g.,



(a) $b = 0.0091$ (b) $b = 0.0396$

Figure 2.4: Shape Parameter $\rho_b(x)$ of Modified Gamma Kernel
Black solid and short-dashed line: $c = 0.6$ and $c = 0.05$ for refined kernel v_{I} (see (2.21)). Gray solid and short-dashed line: $c = 0.6$ and $c = 0.05$ for refined kernel v_{II} (see (2.22)). Black long-dashed line: original modified kernel (see (2.9)). Gray long-dashed line: interior regime of original specification and refined version v_{I}, x/b. Bandwidths of the modified gamma kernel estimator for two DGPs from the simulation study in Section 2.3 are used.



(a) $b = 0.0091$ (b) $b = 0.0396$

Figure 2.5: Gamma Kernel Depending on Shape Parameter
Gamma kernel $K^{\gamma}_{\rho_b(x),b}(u)$ for different values of shape parameter ρ_b. Black solid line: $\rho_b = 0.5$. Black short-dashed line: $\rho_b = 1$. Black long-dashed line: $\rho_b = 1.5$. Gray solid line: $\rho_b = 2$. Bandwidths of the modified gamma kernel estimator for two DGPs from the simulation study in Section 2.3 are used.

the standard gamma kernel estimator, such that

$$\hat{f}_X''^{\gamma}(x) = \frac{1}{n\,b^2} \sum_{i=1}^{n} \frac{\partial^2}{\partial x^2}\; K^{\gamma}_{x/b+1,b}(X_i)\,, \tag{2.25}$$

$$= \frac{1}{n\,b^2} \sum_{i=1}^{n} K^{\gamma}_{x/b+1,b}(X_i)\; \left\{ \left[\ln(X_i/b) - \psi(x/b+1) \right]^2 - \psi_1(x/b+1) \right\},$$

where $\psi(u) := (d/du)\,\ln\Gamma(u)$ and $\psi_1(u) := \left(d^2/du^2\right) \ln\Gamma(u)$ denote the digamma and trigamma function, respectively.

2.3 Simulation Study

In order to obtain a complete picture, we compare basic, modified and refined gamma kernel estimators to standard boundary-corrected versions of the symmetric fixed kernel density estimator (2.1) for a wide range of test densities representing all potential types of shapes near the boundary. Further, this complements simulation studies for the two original gamma kernels put forward in the literature, as e.g., in Chen (2000), who focuses on very specific density settings, and Hagmann and Scaillet (2007), who restrict the range of fixed boundary kernel competitors.

All fixed kernels are of Epanechnikov-type, i.e., $K(u) = 3/4\,\left(1-u^2\right)\,\mathbb{1}_{(-1\le u\le 1)}$, where $\mathbb{1}_{(-1\le u\le 1)}$ denotes an indicator function limiting the support of K to $[-1,1]$. In particular, we report results for the following five competing fixed kernel adjustments.

The reflection estimator proposed by Schuster (1958) is given as

$$\hat{f}_X^{\mathrm{Refl}}(x) = \frac{1}{nb} \sum_{i=1}^{n} K\!\left(\frac{x-X_i}{b}\right) + K\!\left(\frac{x+X_i}{b}\right). \tag{2.26}$$

In the inside of the support for $x \ge 2b$, (2.26) coincides with the standard kernel density estimator $\hat{f}_X^{\mathrm{Fixed}}$ in (2.1).

For the cut-and-normalized estimator $\hat{f}_X^{\mathrm{CaN}}$ introduced by Gasser and Müller (1979), the kernel function K is truncated at $\nu := x/b$ in the boundary region and normalized ensuring integration to unity. Based on the Epanechnikov kernel, it has the form

$$K^{\mathrm{CaN}}(u) = \frac{\left(1-u^2\right)}{\int_{-1}^{\nu}\left(1-u^2\right)du}\,\mathbb{1}_{(-1\le u\le \nu)}. \tag{2.27}$$

General boundary-corrected estimators $\hat{f}_X^{\mathrm{Bound}}$ (see, e.g., Jones, 1993) replace the standard kernel function in the boundary region by a modified version K^{Bound}, which is chosen to meet the following conditions

$$\int_{\nu}^{1} K^{\mathrm{Bound}}(u)\,du = 0, \quad \int_{-1}^{\nu} K^{\mathrm{Bound}}(u)\,du < \infty, \quad \int_{-1}^{\nu} K^{\mathrm{Bound}}(u)\,u\,du = 0. \tag{2.28}$$

Table 2.1: Data Generating Processes for Simulation Study
DGPs are based on i.i.d. samples from different specifications of the generalized F distribution (2.31). We use the following tuples of shape parameters a, m and η. The scale parameter λ is chosen such that the expectation of each DGP is normalized to one. Corresponding shapes of the densities are depicted in Figure 2.6.

DGP	a	m	η
1	1	1	∞
2	0.9	0.7	1.2
3	14	0.2	0.5
4	35	0.08	0.1
5	0.8	2	∞
6	0.55	3	5
7	5	0.3	∞

We use the boundary kernel based on the Epanechnikov kernel, which yields

$$K^{\text{Bound}}(u) = 12\frac{(1+u)}{(1+\nu)^4}\left[\frac{3\nu^2 - 2\nu + 1}{2} + u\,(1-2u)\right] \mathbb{1}_{(-1\leq u\leq \nu)}. \tag{2.29}$$

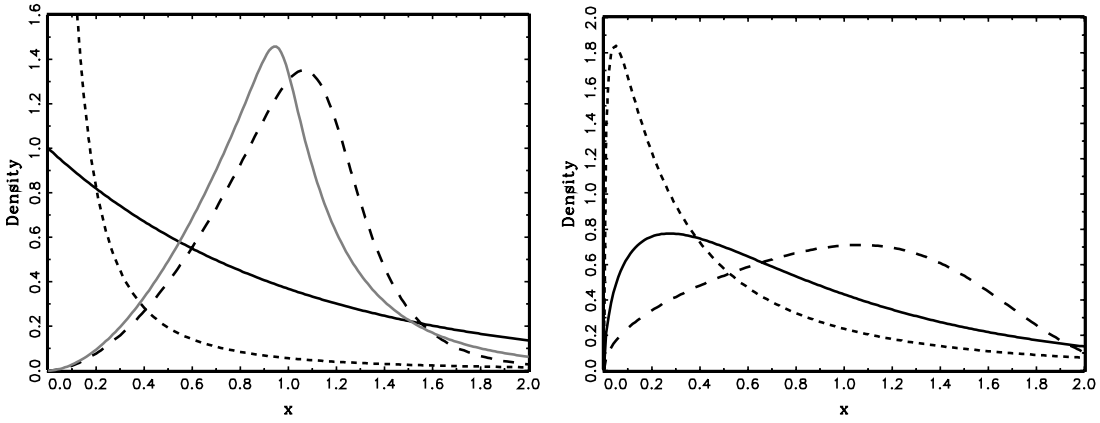
A method that corrects for the possible negativity of the boundary kernel estimates implied by (2.29) was proposed, e.g., by Jones and Foster (1996). The corresponding estimator has the form

$$\hat{f}_X^{\text{JF}}(x) = \hat{f}_X^{\text{CaN}}(x)\,\exp\left\{\frac{\hat{f}_X^{\text{Bound}}(x)}{\hat{f}_X^{\text{CaN}}(x)} - 1\right\}. \tag{2.30}$$

We compare the performance of the above estimators for seven different density functions with nonnegative support, which reflect the variety of practically relevant types of shapes. The densities of DGP 1 and DGP 2 are entirely decreasing and convex with DGP 2 exhibiting pole behavior at zero. The remaining densities are increasing near the boundary. For DGP 3 and 4, the density is locally convex in the boundary region, while for 5,6 and 7, it is concave with a varying degree of steepness. The corresponding density shapes are depicted in Figure 2.6. All DGPs are generated from different specifications of the flexible generalized F distribution, which is based on a gamma mixture of the generalized gamma distribution (see, e.g., Lancaster, 1997). Its marginal density function is given by

$$f_X(x) = \frac{a\,x^{a\,m-1}\,[\eta + (x/\lambda)^a]^{(-\eta-m)}\,\eta^\eta}{\lambda^{a\,m}\,\mathcal{B}(m,\eta)}, \tag{2.31}$$

where $a > 0, m > 0, \eta > 0$ and $\lambda > 0$. $\mathcal{B}(\cdot)$ describes the full beta function with $\mathcal{B}(m,\eta) := \Gamma(m)\,\Gamma(\eta)\,\Gamma(m+\eta)^{-1}$. Table 2.1 shows the values of the shape parameters



(a) DGPs 1 to 4 (b) DGPs 5 to 7

Figure 2.6: Densities Corresponding to Different DGPs
Densities corresponding to tuples of shape parameters in Table 2.1. Left: DGP 1 (black solid), DGP 2 (black short-dashed), DGP 3 (black long-dashed) and DGP 4 (gray solid). Right: DGP 5 (black solid), DGP 6 (black short-dashed) and DGP 7 (black long-dashed).

a, m and η for the seven DGPs considered. To ensure comparability across the different DGPs, the expectation is restricted to one by setting

$$\lambda = \eta^{-1/a} \frac{\Gamma(m)\,\Gamma(\eta)}{\Gamma(m+1/a)\,\Gamma(\eta-1/a)}. \tag{2.32}$$

From each DGP, we draw 1000 random samples $\{X_i\}_{i=1}^n$ of size $n = 400$ and $n = 4000$. To minimize the effects of sampling variation, we follow Zhang (2010) and select the optimal bandwidth for each estimator and DGP by minimizing the integrated mean squared error (IMSE) given as

$$\text{IMSE}\Big\{\hat{f}_X(x)\Big\} = \frac{1}{1000} \sum_{r=1}^{1000} \int_\tau^\infty \Big\{f_X(x) - \hat{f}_X^r(x)\Big\}^2 \, dx, \tag{2.33}$$

where τ is a small number and $\hat{f}_X^r(x)$ denotes the density estimate for the rth simulated sample. Bandwidth selection is conducted using the sample size $n_b = 200$, which requires multiplying the resulting bandwidths by the factor $(n/n_b)^{-1/5}$ for the subsequent analysis. The rescaled bandwidths for $n = 400$ and $n = 4000$ are reported in Table 2.2. The two gamma kernel estimators exhibit noticeably smaller bandwidths in comparison to the other estimators, which can be explained by the reduced variance of the former in the interior part of the support.

Table 2.3 and 2.4 report the IMSEs of the different estimators for the seven DGPs and two samples sizes. IMSEs are computed over the interval $[0, 2]$. For DGPs 3 to 7, we additionally consider shorter intervals that encompass and exclude the mode of the distribution, respectively. Three major results are apparent. First, in a general comparison

Table 2.2: Bandwidths
Bandwidths chosen by minimizing the integrated mean squared error (2.33) using simulated samples with $n = 200$. The following estimators are used. Gam and Gam_m: standard and modified gamma kernel estimator. Fixed: fixed kernel estimator based on the Epanechnikov kernel. Refl: reflection estimator. CaN: cut-and-normalized estimator. Bound: boundary kernel estimator. JF: Jones-Foster estimator.

Est.	**DGP 1**	**DGP 2**	**DGP 3**	**DGP 4**	**DGP 5**	**DGP 6**	**DGP 7**
				n = 400			
Gam	0.0768	0.0042	0.0096	0.0088	0.0571	0.0319	0.0308
Gam_m	0.1163	0.0166	0.0099	0.0091	0.0634	0.0396	0.0336
Fixed	0.1821	0.0176	0.2058	0.1820	0.2678	0.0888	0.4015
Refl	0.4643	0.0391	0.2054	0.1818	0.3569	0.2144	0.3609
CaN	0.4223	0.0307	0.2044	0.1808	0.4308	0.1868	0.3494
Bound	0.7471	0.0086	0.2064	0.1825	0.3824	0.3141	0.4024
JF	0.4223	0.0307	0.2044	0.1808	0.4308	0.1868	0.3494
				n = 4000			
Gam	0.0485	0.0027	0.0061	0.0055	0.0360	0.0201	0.0195
Gam_m	0.0734	0.0104	0.0062	0.0058	0.0400	0.0250	0.0212
Fixed	0.1149	0.0111	0.1299	0.1148	0.1689	0.0561	0.2533
Refl	0.2930	0.0247	0.1296	0.1147	0.2252	0.1353	0.2277
CaN	0.2664	0.0194	0.1290	0.1141	0.2718	0.1179	0.2204
Bound	0.4714	0.0054	0.1302	0.1152	0.2413	0.1982	0.2539
JF	0.2664	0.0194	0.1290	0.1141	0.2718	0.1179	0.2204

Table 2.3: Integrated Mean Squared Errors (n=400)
For DGP 3, 4 and 7, limits x_1 and x_2 are chosen such that $[x_1, x_2]$ encompasses the mode: $x_1 \in \{0.888, 0.600, 0.500\}$ and $x_2 \in \{1.292, 1.090, 1.700\}$. For DGP 5 and 6, $x_1 = x_2$ is chosen such that $[0, x_1]$ includes the mode, where $x_1 = x_2 \in \{0.600, 0.200\}$. The following estimators are used. Gamma and $\text{Gamma}_{\text{mod}}$: standard and modified gamma kernel estimator. Fixed: fixed kernel estimator based on the Epanechnikov kernel. Refl: reflection estimator. CaN: cut-and-normalized estimator. Bound: boundary kernel estimator. JF: Jones-Foster estimator. Results are re-scaled by the factor 10^3.

Estimator	**DGP 1**	**DGP 2**	**DGP 3**	**DGP 4**	**DGP 5**	**DGP 6**	**DGP 7**
				$0 \leq x \leq 2$			
Gamma	4.185	75.151	8.857	9.759	5.828	14.572	5.286
$\text{Gamma}_{\text{mod}}$	3.575	279.733	8.720	9.623	5.422	14.879	4.575
Fixed	17.659	287.118	7.293	8.630	7.182	26.125	3.462
Refl	3.854	320.621	7.309	8.645	9.815	16.738	4.943
CaN	4.171	177.666	7.340	8.670	10.160	17.460	5.308
Bound	3.039	356.335	7.279	8.621	6.651	18.096	3.594
JF	4.259	923.101	7.274	8.612	7.235	17.629	3.792
				$0 \leq x \leq x_1$			
Gamma			2.880	1.729	4.691	11.902	1.804
$\text{Gamma}_{\text{mod}}$			2.553	1.502	4.330	11.590	1.501
Fixed			2.015	1.125	4.991	17.248	0.934
Refl			2.037	1.143	8.285	13.032	2.309
CaN			2.083	1.178	8.956	13.339	2.619
Bound			1.993	1.109	5.250	14.616	1.066
JF			2.017	1.120	6.030	13.508	1.104
				$x_1 \leq x \leq x_2$			
Gamma			4.702	6.416			3.224
$\text{Gamma}_{\text{mod}}$			4.304	6.038			2.549
Fixed			3.788	5.316			2.190
Refl			3.783	5.312			2.303
CaN			3.770	5.295			2.358
Bound			3.796	5.326			2.189
JF			3.770	5.295			2.358
				$x_2 \leq x \leq 2$			
Gamma			1.220	1.573	1.136	2.670	0.258
$\text{Gamma}_{\text{mod}}$			1.806	2.039	1.092	3.289	0.525
Fixed			1.433	2.134	2.190	8.877	0.339
Refl			1.432	2.136	1.530	3.706	0.330
CaN			1.431	2.141	1.204	4.122	0.331
Bound			1.434	2.131	1.401	3.479	0.339
JF			1.431	2.141	1.204	4.122	0.331

Table 2.4: Integrated Mean Squared Errors (n=4000)
For DGP 3, 4 and 7, limits x_1 and x_2 are chosen such that $[x_1, x_2]$ encompasses the mode: $x_1 \in \{0.888, 0.600, 0.500\}$ and $x_2 \in \{1.292, 1.090, 1.700\}$. For DGP 5 and 6, $x_1 = x_2$ is chosen such that $[0, x_1]$ includes the mode, where $x_1 = x_2 \in \{0.600, 0.200\}$. The following estimators are used. Gamma and $\text{Gamma}_{\text{mod}}$: standard and modified gamma kernel estimator. Fixed: fixed kernel estimator based on the Epanechnikov kernel. Refl: reflection estimator. CaN: cut-and-normalized estimator. Bound: boundary kernel estimator. JF: Jones-Foster estimator. Results are re-scaled by the factor 10^3.

Estimator	**DGP 1**	**DGP 2**	**DGP 3**	**DGP 4**	**DGP 5**	**DGP 6**	**DGP 7**
				$0 \leq x \leq 2$			
Gamma	1.011	21.601	2.246	2.758	2.069	6.924	1.218
$\text{Gamma}_{\text{mod}}$	0.679	161.086	2.237	2.760	1.734	7.307	0.986
Fixed	8.122	46.145	1.331	1.683	1.960	8.504	0.734
Refl	0.803	154.026	1.332	1.684	4.415	10.797	1.203
CaN	0.890	56.600	1.334	1.682	5.443	10.075	1.318
Bound	0.471	62.457	1.329	1.684	2.106	10.593	0.684
JF	0.585	616.087	1.327	1.676	2.663	8.142	0.721
				$0 \leq x \leq x_1$			
Gamma			0.450	0.271	1.831	6.484	0.322
$\text{Gamma}_{\text{mod}}$			0.349	0.207	1.532	6.567	0.216
Fixed			0.360	0.193	1.592	7.029	0.267
Refl			0.362	0.194	4.149	10.148	0.723
CaN			0.367	0.198	5.227	9.364	0.830
Bound			0.357	0.191	1.859	9.886	0.216
JF			0.360	0.192	2.447	7.431	0.234
				$x_1 \leq x \leq x_2$			
Gamma			1.495	2.200			0.826
$\text{Gamma}_{\text{mod}}$			1.337	2.055			0.567
Fixed			0.721	1.118			0.405
Refl			0.720	1.116			0.421
CaN			0.717	1.110			0.430
Bound			0.722	1.121			0.405
JF			0.717	1.110			0.430
				$x_2 \leq x \leq 2$			
Gamma			0.293	0.281	0.238	0.440	0.071
$\text{Gamma}_{\text{mod}}$			0.542	0.488	0.202	0.741	0.202
Fixed			0.239	0.361	0.369	1.475	0.062
Refl			0.239	0.361	0.266	0.650	0.058
CaN			0.239	0.362	0.216	0.711	0.058
Bound			0.239	0.361	0.247	0.706	0.062
JF			0.239	0.362	0.216	0.711	0.058

with the standard fixed kernel adjustments, gamma kernel estimators appear to offer a satisfactory performance. They are clearly more precise for DGPs 2, 5 and 6, while yielding similar (or only slightly higher) IMSEs in the remaining cases. In particular, the single largest improvement in favor of the (standard) gamma kernel is achieved in the pole scenario of DGP 2. Note that when the applied polynomial transformation for the method of Marron and Ruppert (1994) was close to the true pole behavior, we could also construct a fixed kernel estimator with a similar or even better precision for DGP 2. Corresponding results, however, were not robust to deviations of the transformation from the true density shape near zero, implying a high risk of extremely large IMSEs in practice. Due to the tailored construction of the above method for pole situations only, the IMSE records for any other form of the density were largely inferior to the rest. Therefore, we do not report results for this estimator.

Second, the simulation evidence confirms the relationship between the relative performance of the standard and modified gamma kernel estimator and the shape of the underlying density. If the latter has first and second derivatives of opposing sign in the interior of the support, as is the case for DGPs 3, 4, 6 and 7 in the subinterval to the right of the mode, the standard gamma kernel yields noticeably lower IMSEs (see bottom panel). When considering the entire interval $[0, 2]$, the basic gamma kernel is more precise for DGPs 2 and 6 with the most striking gains occurring in the former scenario, as it corresponds to a globally convex density with pole at zero. Finally, the above relation breaks down within the boundary region due to the involvement of the factor $\xi_b(x)$ in the asymptotic bias (see (2.10)). For DGPs 5 and 6, the modified gamma kernel implies lower IMSEs over the leftmost subinterval, in which the corresponding densities are increasing and concave (see lower top panel).

The simulation results stress the importance of determining pole situations in advance, which can be achieved by examining the normalized density derivative $D(x)$ in the boundary region. We estimate the latter as in (2.19) using the modified gamma kernel for the points $x \in \{0, b, 2b\}$, where b is the bandwidth of the corresponding estimator. Table 2.5 reports descriptive statistics of the estimates for $n = 400$. In case of DGP 2, these estimates are highly negative at all three points, demonstrating that our simple method is able to detect a pole at zero. We obtain negative estimates at all or at distinct points also for DGPs 1 and 6, but their magnitude is considerably lower than in the above true pole scenario.

As was argued in Section 2.2.2, whenever no pole situation is detected, the modified gamma kernel in its original or refined form should be used. The IMSEs of the three corresponding estimators are displayed in Table 2.6. For the refined kernels v_{I} and v_{II}, a set of values for the modification parameter c is considered. To ensure comparability, we apply the bandwidths b of the original modified gamma kernel to all estimators and also use $2b$ as the upper integration limit in the IMSE calculations. The main finding is that the refined kernel v_{I} exhibits a high precision in all situations for which the modified kernel should be considered, i.e., all DGPs except the second one. The improvement with respect to the original specification is particularly pronounced, accompanied by low optimal values of the parameter c, in case of densities with concave shape near the

Table 2.5: Summary Statistics of Normalized Density Derivative
Descriptives for estimates of the ratio $D(x) := f'_X(x)/f_X(x)$ based on the modified gamma kernel. The estimator (2.19) is used. $n = 400$.

x	**DGP 1**	**DGP 2**	**DGP 3**	**DGP 4**	**DGP 5**	**DGP 6**	**DGP 7**
				Mean			
0	-0.226	-13.982	73.923	80.433	1.899	1.265	4.710
b	-0.711	-27.341	182.844	205.189	3.260	0.756	8.741
$2b$	-0.941	-23.570	185.301	198.408	1.892	-1.558	6.763
				1st Quartile			
0	-0.399	-15.110	67.995	73.900	1.473	0.737	2.707
b	-1.017	-29.185	165.094	184.470	2.467	-0.257	4.562
$2b$	-1.180	-25.188	161.913	172.373	1.247	-2.386	3.800
				Median			
0	-0.220	-13.928	75.557	82.861	1.922	1.284	4.890
b	-0.727	-27.382	187.737	212.998	3.209	0.710	8.395
$2b$	-0.939	-23.557	191.768	208.193	1.865	-1.568	6.629
				3rd Quartile			
0	-0.059	-12.794	81.790	89.863	2.350	1.798	7.020
b	-0.421	-25.602	206.611	235.238	4.025	1.764	12.854
$2b$	-0.703	-21.906	216.830	236.042	2.493	-0.786	9.496

boundary, as in DGPs 5,6 and 7. Further, the refined kernel v_{II} is at roughly the same level as the traditional parameterization and even yields the lowest IMSE for DGP 1 when $n = 400$. However, recall that this specification makes the boundary region smaller and has neither its mean nor mode at the point of estimation for $x > 2bc$ (see Section 2.2.3). These properties cause a vastly lower precision compared to the other specifications in the interior part of the support. Corresponding simulation results are available upon request.

Finally, Table 2.6 shows that the performance of the refined modified gamma kernel estimators is highly dependent on the value of the modification parameter c. This is underlined by Figure 2.7, which depicts plots of the root mean squared errors (RMSEs) of the estimators based on the original modified gamma kernel and the refined version v_{I} for several values of c. The plots also illustrate that the choice of c determines for which part of the support the original estimator can or cannot be outperformed. E.g. in case of DGP 4, specification v_{I} almost consistently exhibits lower RMSEs for $c = 0.6$ or $c = 0.1$, while providing precise estimates only in a small neighborhood of $x = 0$ if $c = 0.01$.

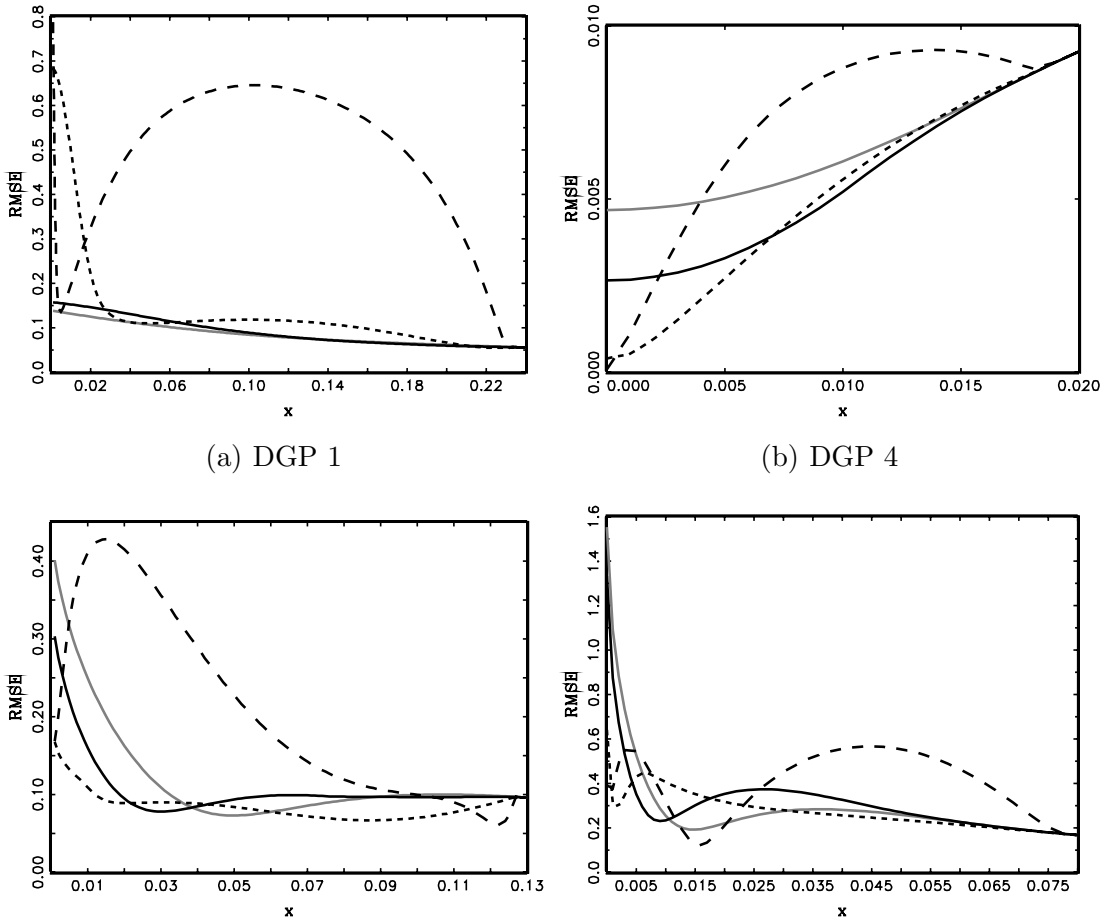
Since in practice, the modification parameter c has to be chosen ex-ante, we examine how well the data-driven method introduced in Section 2.2.3 can "track" the optimal values according to Table 2.6. We estimate the unknown quantities entering the objective function (2.24) as was outlined above. Figure 2.8 displays means, medians and quartiles of the resulting estimates of the (transformed) objective function $\mathcal{Q}(c) := \mathcal{M}(2bc)$, where b is the bandwidth of the modified gamma kernel. A comparison with the IMSEs from Table 2.6 shows that, for DGPs 5, 6 and 7, the means, in particular, have local minima close to the values of c yielding the lowest IMSEs of the estimator based on the refined modified kernel v_{I}. For DGP 4, finding a unique minimum is more difficult, which corresponds to the fact that several values of c imply equal IMSEs. These results suggest that if suitable starting values are chosen, the above approach can determine the optimal value of c with reasonable precision.

2.4 Application: Intraday Trading Volumes and Return Volatility

To demonstrate the practical relevance of the above methodology, we employ the latter to compute semiparametric estimates of the conditional distributions of high-frequency trading volumes and return volatilities of stocks traded at the New York Stock Exchange (NYSE). Modeling high-frequency trading volumes is, for instance, relevant for trading strategies replicating the (daily) volume weighted average price (VWAP). Estimates of conditional volatility distributions are crucial for the pricing of volatility derivatives. Examples include options and futures on the CBOE Volatility Index (VIX) trading at the Chicago Board Options Exchange (CBOE).

Table 2.6: Integrated MSE for Refined Modified Gamma KDE v_{I} & v_{II}
Refined modified gamma kernel estimators as defined in (2.8) and (2.21) or (2.22). $c = 1^*$ denotes original modified gamma kernel from (2.8) and (2.9). IMSEs are computed from 0 to $2b$. Bandwidths of the original modified gamma kernel are used. Results for $n = 400$ and $n = 4000$ are rescaled by the factor 10^4 and 10^5, respectively.

	c	DGP 1	DGP 2	DGP 3	DGP 4	DGP 5	DGP 6	DGP 7
					n = 400			
	1^*	18.230	1971.501	0.005	0.009	24.452	99.402	2.713
v_{I}	0.9	18.389	2238.201	0.005	0.008	22.112	98.176	2.555
	0.8	18.927	3280.673	0.005	0.008	19.936	97.139	2.416
	0.7	19.976	6045.533	0.004	0.007	17.940	96.462	2.299
	0.6	21.776	13836.526	0.004	0.007	16.123	96.310	2.201
	0.3	39.301	16309119	0.004	0.007	11.443	96.476	1.991
	0.1	74.542	3193545	0.004	0.007	9.150	69.093	1.872
	0.05	99.107	29681669	0.004	0.007	18.179	45.087	2.162
	0.01	623.499	3640.61	0.008	0.012	72.511	144.333	10.431
v_{II}	0.9	17.878	1453.286	0.006	0.010	23.301	97.185	2.722
	0.8	18.157	1040.206	0.007	0.011	22.507	95.373	2.780
	0.7	19.768	837.209	0.008	0.013	22.470	94.715	2.937
	0.6	24.050	996.348	0.009	0.015	23.755	96.878	3.285
	0.3	104.500	6653.866	0.027	0.042	43.068	195.018	8.734
	0.1	629.167	24190.131	0.364	0.530	81.172	874.974	48.989
	0.05	1115.938	32319.507	3.483	5.224	170.476	1460.694	105.800
	0.01	1688.05	38771.282	69.245	70.217	321.568	2185.416	48.598
					n = 4000			
	1^*	25.397	10185.573	0.002	0.002	94.207	594.231	5.050
v_{I}	0.9	23.982	16235.410	0.002	0.002	80.850	562.997	4.439
	0.8	23.446	29262.905	0.002	0.001	68.844	533.361	3.912
	0.7	24.057	55516.877	0.001	0.001	58.242	506.235	3.472
	0.6	26.498	110274.880	0.001	0.001	49.013	483.124	3.115
	0.3	109.444	16058082	0.001	0.001	27.206	431.343	2.493
	0.1	721.532	561573.060	0.001	0.001	11.743	283.302	2.029
	0.05	740.051	57227048	0.001	0.001	31.993	131.333	2.011
	0.01	1716.071	6091.886	0.002	0.002	523.996	244.364	16.592
v_{II}	0.9	25.709	5653.425	0.002	0.002	88.817	571.585	5.026
	0.8	27.903	2185.219	0.002	0.002	86.931	550.769	5.212
	0.7	33.997	682.415	0.003	0.003	91.895	534.382	5.827
	0.6	47.993	2474.760	0.003	0.003	108.945	526.934	7.271
	0.3	306.934	53804.758	0.009	0.011	334.800	750.698	31.256
	0.1	2502.053	207647.900	0.145	0.209	634.147	3412.288	212.589
	0.05	5517.404	277217.360	1.424	2.168	757.340	6670.104	540.434
	0.01	10682.698	335603.280	373.201	424.930	1566.361	12161.535	433.552



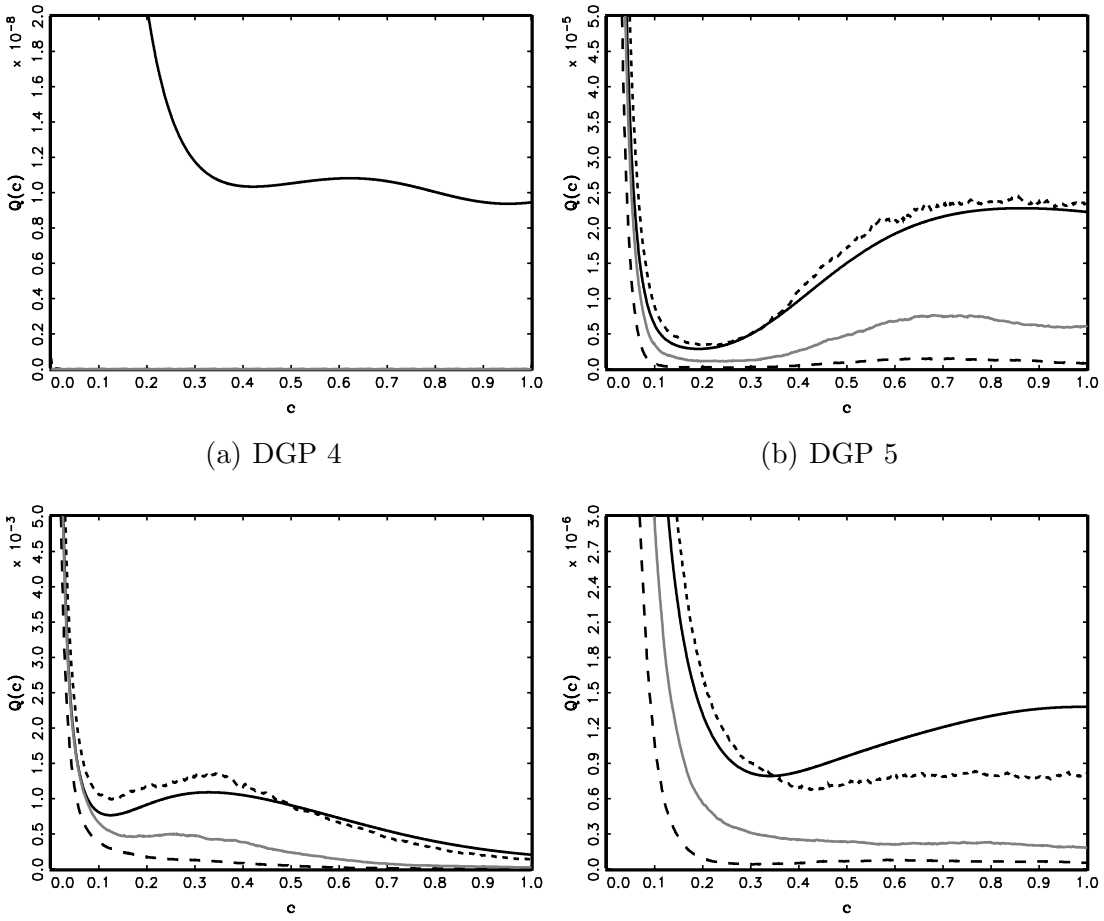
(a) DGP 1

(b) DGP 4

(c) DGP 5

(d) DGP 6

Figure 2.7: RMSE of Refined Modified Gamma KDE v_{I}
Refined modified gamma kernel v_{I} as defined in (2.8) and (2.21). Black solid line: $c = 0.6$. Black short-dashed line: $c = 0.1$. Black long-dashed line: $c = 0.01$. Gray solid line: $c = 1^*$ (original modified gamma kernel). $n = 400$. Bandwidths of the original modified gamma kernel are used.



(a) DGP 4

(b) DGP 5

(c) DGP 6

(d) DGP 7

Figure 2.8: Objective Function for Choice of c
Mean (black solid), median (gray solid), first (black long-dashed) and third (black short-dashed) quartile of (transformed) objective function for choice of the modification parameter c in the refined modified gamma kernel v_{I} as defined in (2.8) and (2.21). The transformed objective function is $\mathcal{Q}(c) := \mathcal{M}(2bc)$, where $\mathcal{M}(x)$ is given in (2.24) and b denotes the bandwidth of the original modified gamma kernel. $n = 400$.

2.4.1 Modeling Intraday Trading Volumes

We consider transaction data for Citigroup from the last trading week of February 2009. The raw sample is filtered by deleting transactions that occurred outside regular trading hours from 9:30 am to 4:00 pm, computing cumulated trading volumes over 15-second intervals and removing zero observations, which yields a sample size of 7452.[2] To capture the well-known intraday seasonalities of high-frequency trading variables (see, e.g., Hautsch (2004) for an overview), we divide the cumulated volumes by a seasonality component, which is pre-estimated employing a cubic spline function.

An important property of the resulting (deseasonalized) trading volumes is the strong persistence, as evidenced by the highly significant Ljung-Box statistics in Table 2.7. The most widely-used parametric framework for this type of data, see, e.g., Brownlees et al. (2010), is the multiplicative error model (MEM) originally proposed by Engle (2002b). Accordingly, we decompose the t-th trading volume, $x_t^{(v)}$, as

$$x_t^{(v)} = \mu_t^{(v)} \varepsilon_t^{(v)}, \quad \varepsilon_t^{(v)} \sim \text{i.i.d. } \mathcal{D}(1)\,, \tag{2.34}$$

where $\mu_t^{(v)}$ denotes the conditional mean given the past information set $\mathcal{F}_{t-1}^{(v)}$ and is assumed to evolve according to the dynamics described in Appendix B.1. $\varepsilon_t^{(v)}$ is a disturbance following an unspecified distribution $\mathcal{D}(1)$ with positive support and $E\big[\varepsilon_t^{(v)}\big] = 1$. Assuming MEM-type dynamics would allow to apply gamma kernel estimators to trading volumes directly and estimate their unconditional density $f_X\big(x_t^{(v)}\big)$ consistently (see Bouezmarni and Rombouts, 2010). Our object of interest, the conditional density given the past information set $\mathcal{F}_{t-1}^{(v)}$, can be estimated semiparametrically in a straightforward way, as the MEM structure implies the basic relationship

$$f_X\big(x_t^{(v)}|\mathcal{F}_{t-1}^{(v)}\big) = f_\varepsilon\big(x_t^{(v)}/\mu_t^{(v)}\big)/\mu_t^{(v)}. \tag{2.35}$$

We consider a two-step approach. First, we estimate $\mu_t^{(v)}$ by exponential QML and generate residuals $\hat{\varepsilon}_t^{(v)} := x_t^{(v)}/\hat{\mu}_t^{(v)}$, which are consistent estimates of the i.i.d errors $\varepsilon_t^{(v)}$ (see, e.g., Drost and Werker, 2004). Second, we estimate $f_\varepsilon\big(x_t^{(v)}/\mu_t^{(v)}\big)$ nonparametrically employing gamma kernels. The consistency and parametric rate of convergence of the conditional mean estimates enable us to use the MEM residuals as inputs without affecting the asymptotics of the kernel density estimators.

Nonparametric estimation of the error density requires the choice of the appropriate type of gamma kernel, i.e., standard or modified in the original and refined version (specification v_{I}). To ensure comparability and boundary regions of equal size, we consider the least-squares cross-validation (LSCV) bandwidth of the standard gamma kernel estimator in all cases. In particular, we use the bandwidth b^* that minimizes a nearly

[2]For a detailed discussion of the treatment of zero observations in the context of financial high-frequency data, see Chapter 1.

Table 2.7: Ljung-Box Statistics for Trading Volumes and Realized Kernel Estimates
$Q(l)$: Ljung-Box statistic associated with l lags. The 5% (1%) critical values associated with lag lengths 20, 50 and 100 are 31.41 (37.57), 67.51 (76.15) and 124.34 (135.81). We consider deseasonalized nonzero 15-second trading volumes of Citigroup and realized kernel (RK) estimates for JP Morgan.

	Volume	**RK**
$Q(20)$	10349.281	5045.309
$Q(50)$	19447.096	9834.944
$Q(100)$	31353.699	14012.591

unbiased estimate of the integrated mean squared error, i.e.,

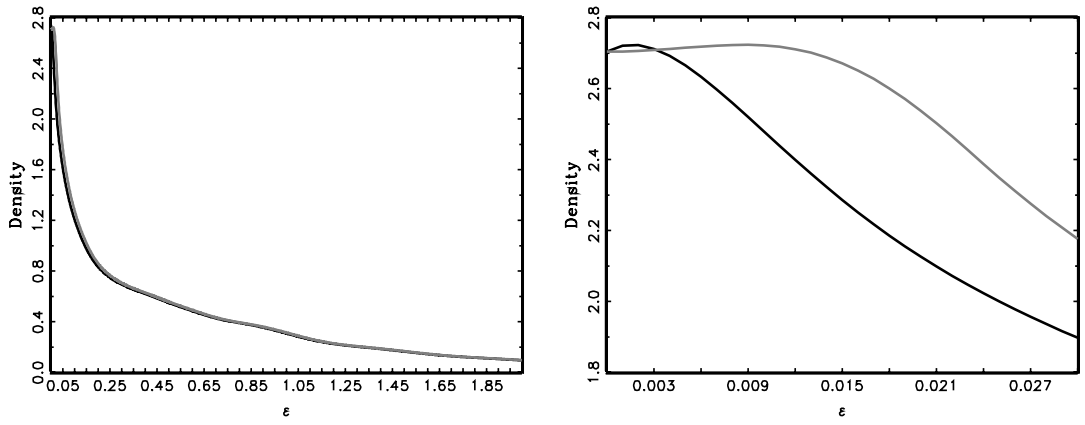
$$CV(b) := \frac{1}{n^2} \sum_i \sum_j \int_\tau^\infty K^\gamma_{x/b+1,b}\left(\hat{\varepsilon}_i^{(v)}\right) K^\gamma_{x/b+1,b}\left(\hat{\varepsilon}_j^{(v)}\right) dx \qquad (2.36)$$
$$- \frac{2}{n\,(n-1)} \sum_i \sum_{j \neq i} K^\gamma_{x_i/b+1,b}\left(\hat{\varepsilon}_j^{(v)}\right),$$

which yields the bandwidth $b^* = 0.0118$. See Hjort and Glad (1995) for details on (nearly) unbiased cross-validation. Further, we estimate the normalized density derivative $D\big(\varepsilon_t^{(v)}\big)$ for $\varepsilon_t^{(v)} \in \{0, b^*, 2b^*\}$ as in (2.19) based on the modified gamma kernel. The corresponding results in Table 2.8 show that two out of three estimates are considerably negative, which indicates a possible pole situation and suggests the use of the standard gamma kernel. Figure 2.9 displays estimates of the error density $f_\varepsilon\big(\varepsilon_t^{(v)}\big)$ based on the standard and, for comparison, modified gamma kernel for the boundary region and a larger part of the support. While for both density estimates, the probability mass is quite concentrated close to the origin, the standard gamma kernel, being the method of choice, yields an estimate that lies clearly below the density implied by the modified kernel for the major part of the boundary region.

Finally, Figure 2.10 shows estimates of the conditional density of trading volumes for February 26 and 27, 2009, at 11am EST. On the latter day, Citigroup announced that the US treasury would be taking a major equity stake in the company, while the former day is included for comparison. As an alternative to the semiparametric approach, the plot also features the conditional density implied by maximum likelihood estimates of the MEM (2.34) assuming that the errors follow the widely-used gamma distribution (e.g. Engle and Gallo, 2006). The impact of the announcement on trading activity related to the Citigroup stock is clearly visible, as the conditional volume distribution for February 27 assigns considerably less weight to small transactions. The semiparametric density estimates and their parametric counterparts are quite close to each other in the interior of the support. The major difference occurs at the origin where the parametric densities exhibit a pole, which is not the case for the semiparametric estimates.

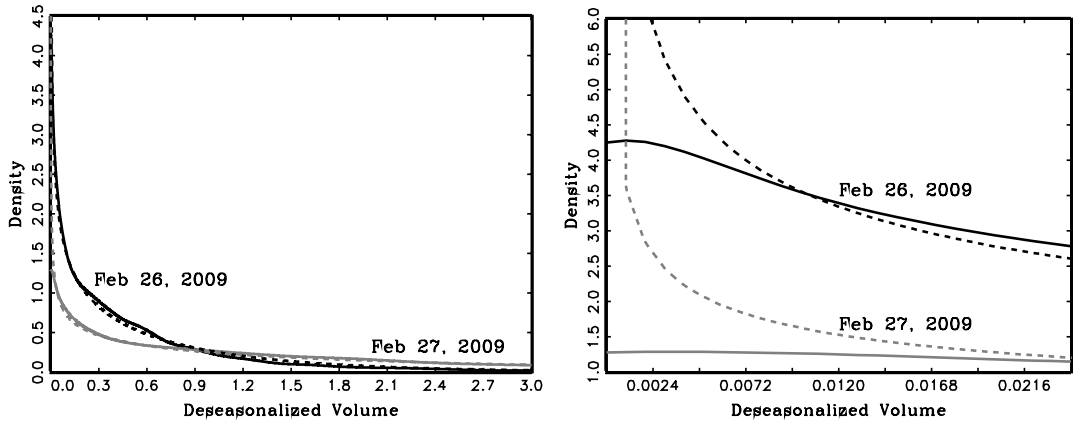
Table 2.8: Estimates of Normalized Density Derivative for MEM Errors
Estimates of the ratio $D(\varepsilon_t^{(m)}) := f'_\varepsilon(\varepsilon_t^{(m)})/f_\varepsilon(\varepsilon_t^{(m)})$, $m = v, rk$, based on the modified gamma kernel in the boundary region as in (2.19). $\varepsilon_t^{(m)}$ are errors from the MEM structure (2.34) fitted to deseasonalized nonzero 15-second trading volumes of Citigroup and realized kernel (RK) estimates for JP Morgan. b^* is LSCV bandwidth of the standard gamma kernel estimator: 0.0118 for trading volumes and 0.0206 for realized kernel estimates.

$\varepsilon_t^{(m)}$	**Volume**	**RK**
0	0.293	26.283
b^*	-10.100	59.915
$2b^*$	-14.399	53.235



(a) Full Support (b) Boundary

Figure 2.9: Estimates of MEM Error Density for Intraday Trading Volumes
Estimates of the density $f_\varepsilon(\varepsilon_t^{(v)})$ from the MEM structure (2.34) fitted to deseasonalized nonzero 15-second trading volumes of Citigroup. Black solid line: standard gamma kernel. Gray solid line: modified gamma kernel. LSCV bandwidth of the standard gamma kernel, $b^* = 0.0118$, is used for both estimators.



(a) Full Support (b) Boundary

Figure 2.10: (Semi-)Parametric Conditional Density of Intraday Trading Volumes
Conditional densities of deseasonalized nonzero 15-second trading volumes of Citigroup at time t given past information $\mathcal{F}_{t-1}^{(v)}$. Based on the MEM structure (2.34) and the relationship (2.35). Parametric estimates (dashed lines) are implied by a ML approach assuming gamma-distributed errors $\varepsilon_t^{(v)}$. Semiparametric estimates (solid lines) rely on QML estimates of $\mu_t^{(v)}$ and nonparametric estimates of $f_\varepsilon\big(x_t^{(v)}/\mu_t^{(v)}\big)$ using the standard gamma kernel. Conditional densities are estimated for 11am EST on February 26 (black lines) and February 27, 2009 (gray lines).

2.4.2 Modeling Realized Volatility

Realized volatility measures computed from high-frequency data allow to construct more accurate estimates of the underlying lower frequency volatility (see, e.g., Andersen et al., 2010a). We employ mid-quotes for JP Morgan from January 2006 to December 2009, which corresponds to 983 trading days, and clean the raw data as suggested in Barndorff-Nielsen et al. (2008b). The realized volatility for day t is simply defined as the sum of squared (mid-quote) returns $r_{i,t}$, $i = 1, \ldots, N_t$. Barndorff-Nielsen and Shephard (2002) show that, in the absence of noise and with the number of intraday returns approaching infinity, this basic estimator is consistent for the latent integrated volatility, which under regularity conditions, provides an unbiased measure of the conditional variance of (daily) returns. In practice, observed prices are contaminated by microstructure effects, causing an inconsistency of the basic realized volatility estimator (e.g. Hansen and Lunde, 2006). Hence, we consider the noise-robust realized kernel estimator, which was proposed by Barndorff-Nielsen et al. (2008a) and takes the form

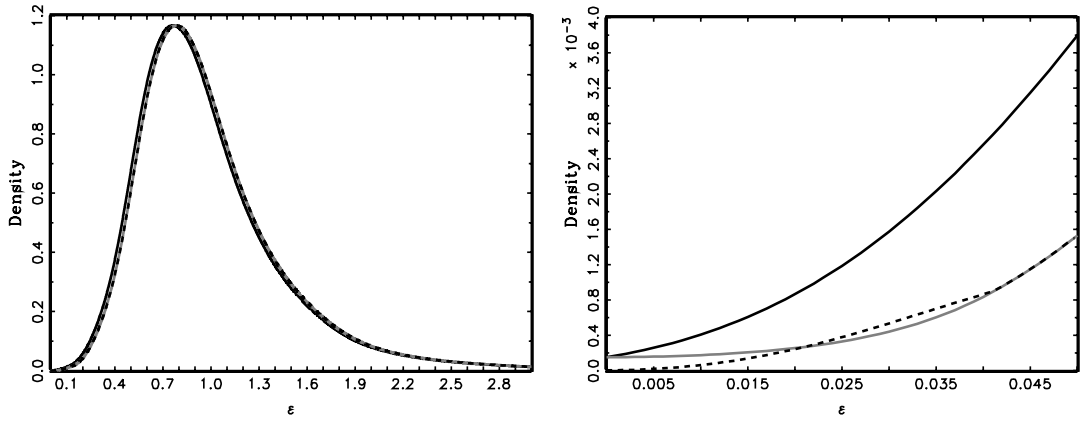
$$x_t^{(rk)} := \gamma_0 + \sum_{h=1}^{H} k\left(\frac{h-1}{H}\right)(\gamma_h + \gamma_{-h}), \qquad \gamma_h := \sum_{i=1}^{n_t} r_{i,t}\, r_{i-h,t}, \tag{2.37}$$

where $k(\cdot)$ is the Parzen kernel and H the bandwidth.[3] Since (filtered) realized kernel estimates are used as inputs for kernel density estimators below, the two bandwidths involved have to be balanced in a way similar to Corradi et al. (2009) who propose nonparametric conditional density estimators for the integrated volatility. We ensure that their assumption A.1 is met by choosing H as in Section 4.3 of Barndorff-Nielsen et al. (2008a).[4]

Table 2.7 shows that the realized kernel estimates exhibit a similar persistence as trading volumes, which we account for by following Engle and Gallo (2006) and imposing a flexible MEM structure. Hence, we model the realized kernel value for day t, $x_t^{(rk)}$, analogously to (2.34), where the assumptions for the errors $\varepsilon_t^{(rk)}$ remain the same, while a slightly different specification is chosen for the conditional mean $\mu_t^{(rk)}$ (see Appendix B.1). We compute semiparametric estimates of the conditional density $f_X\big(x_t^{(rk)}|\mathcal{F}_{t-1}^{(rk)}\big)$ using the same approach as in Section 2.4.1, which in the given application, can be considered as a simple alternative to the fully nonparametric procedure proposed in Corradi et al. (2009). As Table 2.8 reports, the estimates of the normalized density derivative for the MEM errors are consistently positive, indicating that the corresponding density should be estimated using a modified gamma kernel. Thus, we first determine the optimal value of the modification parameter c for the refined specification v_{I} by minimizing the objective function (2.24). We compute the required pilot estimates of the unknown density and its first two derivatives as outlined in Section 2.2.3, which yields the parameter value $c^* = 0.0863$.

[3] The number of returns used for the computation of the realized kernel, n_t, is lower than the total number of observations N_t due to the so-called jittering procedure. See Barndorff-Nielsen et al. (2008a) for details.

[4] To estimate the so-called noise-to-signal ratio, we follow Barndorff-Nielsen et al. (2008b).

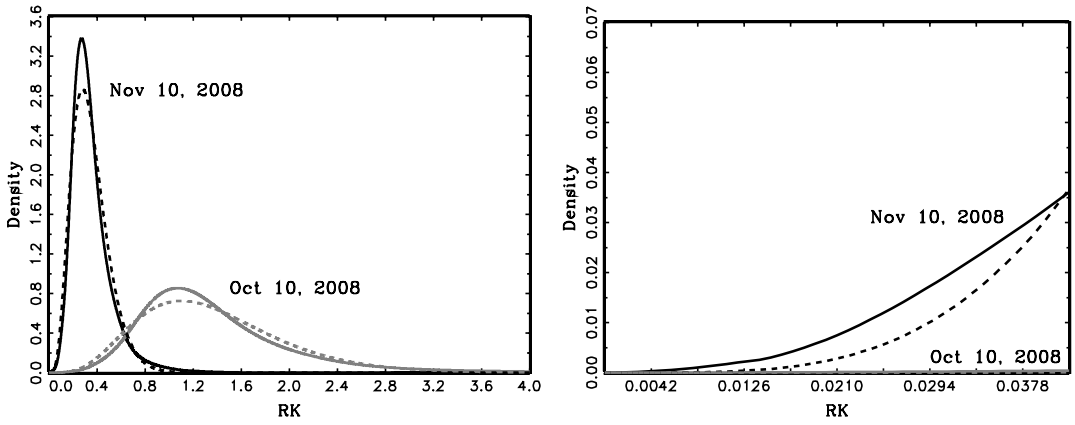


(a) Full Support (b) Boundary

Figure 2.11: Estimates of MEM Error Density for Realized Kernel Estimates
Estimates of the density $f_\varepsilon\big(\varepsilon_t^{(rk)}\big)$ from the MEM structure (2.34) fitted to realized kernel estimates for JP Morgan. Black solid line: standard gamma kernel. Gray solid line: modified gamma kernel. Black dashed line: refined modified gamma kernel v_{I}. LSCV bandwidth of the standard gamma kernel, $b^* = 0.0206$, is used for all estimators.

Estimates of the MEM error density implied by all three types of gamma kernels considered are displayed in Figure 2.11 and indicate the following major results. First, as compared to the error density based on trading volumes in Figure 2.9, the mode of the distribution is further to the interior of the support. Second, the density exhibits a similar degree of right-skewness as was reported for the unconditional distribution of realized volatilities by Andersen et al. (2001). Finally, the density estimate based on the refined modified kernel tends to zero when approaching the boundary instead of taking a strictly positive value at $\varepsilon_t^{(rk)} = 0$. This effect is caused by the low value of the modification parameter c, which pushes the shape parameter $\rho_b^{v_{\mathrm{I}}}\big(\varepsilon_t^{(rk)}\big)$ below one when smoothing at the boundary (see (2.21)). A distribution of stock return volatility with vanishing probability mass close to the boundary is in line with financial theory, since stocks are "risky" assets for which investors demand a volatility premium (see, e.g., Merton, 1973).

Figure 2.12 displays conditional density estimates of realized kernel values for two days during the financial crisis 2007 – 2008: October 10, 2008, when the DJIA index fell by 8% at the start of the trading day, and November 10, 2008, when a major restructuring of the AIG bailout plan was announced. The density estimates are based on our semiparametric procedure using the refined modified gamma kernel and the parametric approach from Section 2.4.1. Except for some discrepancies around the mode and in the boundary region, the parametric estimates roughly match the semiparametric ones, indicating that the gamma distribution is a reasonable assumption for the MEM errors. With respect to dynamic changes, the conditional densities reflect the more unstable



(a) Full Support (b) Boundary

Figure 2.12: (Semi-)Parametric Conditional Density of Realized Kernel Estimates
Conditional densities of realized kernel estimates for JP Morgan at day t given past information $\mathcal{F}_{t-1}^{(rk)}$. Based on the MEM structure (2.34) and the relationship (2.35). Parametric estimates (dashed lines) are implied by a ML approach assuming gamma-distributed errors $\varepsilon_t^{(rk)}$. Semiparametric estimates (solid lines) rely on QML estimates of $\mu_t^{(rk)}$ and nonparametric estimates of $f_\varepsilon\big(x_t^{(rk)}/\mu_t^{(rk)}\big)$ using the refined modified gamma kernel v_{I}. Conditional densities are estimated for October 10 (gray lines) and November 10, 2008 (black lines). Realized kernel estimates are annualized.

market environment on October 10, since the corresponding volatility distribution has its mode further away from the origin and is more dispersed. Further, as in case of the unconditional error density, the probability mass is vanishing close to the boundary for both days and estimators considered.

2.5 Conclusion

Gamma kernel estimators vary their shape according to the point of estimation along the support. For positive-valued random variables, this location adaptiveness avoids the boundary bias associated with standard fixed kernel estimators, while yielding strictly nonnegative density estimates by construction. We show for various density shapes that, in finite samples, the two original gamma kernel estimators outperform all boundary and boundary-corrected fixed kernel-type estimators in the boundary region, especially for settings with a pronounced probability mass close to zero. For all other setups and in the interior of the support, their finite sample performance is comparable to the one of fixed-type boundary kernels. Moreover, based on asymptotic considerations and finite sample illustrations, we find that, for pole situations at zero, the two gamma kernel estimators differ substantially. In fact, the standard type is superior to the generally

used modified version in this case. We therefore suggest a simple criterion to check for such situations. For all other settings, we propose a refined modified version of the gamma kernel estimator, which further improves upon the performance of the original modified kernel. Our technique is complemented by a data-driven approach for choosing the modification parameter in the new refined gamma kernel. In two application settings, we demonstrate that, in particular in high-frequency finance, the suggested methodology yields superior results of practical impact.

Chapter 3

Do High-Frequency Data Improve High-Dimensional Portfolio Allocations?

This chapter is based on Hautsch, Kyj, and Malec (2011).

3.1 Introduction

With the rise in mutual fund and exchange-traded fund (ETF) investing, quantitative short-term management of vast portfolios has emerged as a topic of great interest. For allocation decisions, forecasts of high-dimensional covariance matrices constitute a crucial input, which initiated a body of literature on the performance of various methods based on asset return data measured up to a daily frequency (see, e.g., Chan et al., 1999; Jagannathan and Ma, 2003). Although the work of Andersen et al. (2001), Barndorff-Nielsen and Shephard (2004) and Barndorff-Nielsen et al. (2011), among others, opened up a new channel for increasing the precision of covariance matrix estimates and forecasts by exploiting high-frequency (HF) data, existing empirical studies examine its benefits for portfolio selection only in moderate dimensions (e.g. Fleming et al., 2003; Liu, 2009). This chapter evaluates the potential of HF data for portfolio selection in a realistic high-dimensional framework.

While ensuring a high precision, we face major technical and practical challenges when constructing covariance matrix forecasts for vast-dimensional portfolio applications. First, forecasts have to be both positive definite and well-conditioned. These properties can be guaranteed by having sufficiently long estimation windows, sampling frequently enough within a fixed window, imposing a parametric specification or applying suitable regularization techniques. The latter include factor structures, e.g., based on principal components, methods from random matrix theory, such as eigenvalue cleaning (see Laloux et al., 1999), or shrinkage techniques as proposed in Ledoit and Wolf (2003). Second, covariance matrix predictions have to balance responsiveness (to new

information) and a certain degree of stability. The latter property is crucial for preventing high transaction costs caused by excessive portfolio rebalancing and can be ensured by smoothing the estimates appropriately.

Motivated by these requirements, we address the following research questions: (i) Do HF-based forecasts generally outperform low-frequency-based approaches and – if yes – over which time horizons? (ii) Which regularization methods are (empirically) superior? (iii) How important is it to smooth estimates over time? (iv) How well do naive predictions of covariance matrices (i.e., random walk forecasts) perform compared to corresponding dynamic forecasting models? (v) How do results change in dependence of the dimension of the underlying portfolio?

We answer these questions in an extensive and thorough empirical study by focusing on the problem of constructing global minimum variance (GMV) portfolios based on the constituents of the S&P 500 index over a four-year period covering the 2008 financial crisis. Studying *global* minimum variance portfolios (in contrast to minimum variance portfolios for a given expected return) has the important advantage that the corresponding weights are determined solely by forecasts of the conditional covariance matrices over the given investment horizon. This property is tantamount to pure volatility timing strategies and avoids the inherent noisiness of conditional mean predictions, overshadowing the analysis and blurring the role of *covariance* forecasts, (see, e.g., Jagannathan and Ma, 2003). We obtain HF-based covariance matrix estimates by applying the blocked realized kernel (BRK) by Hautsch et al. (2012) to mid-quote data. These estimates are smoothed over different time windows, regularized by eigenvalue cleaning or imposing a factor structure and, finally, utilized to construct both naive predictions and forecasts based on a simple dynamic specification. We benchmark the HF forecasts with prevailing approaches employing *daily* returns. In particular, we use multivariate GARCH models, rolling window sample covariance matrices regularized in different ways, as well as both classic and state-of-the-art RiskMetrics approaches. The competing methods are evaluated in terms of the (estimated) conditional portfolio volatility and important characteristics of the implied portfolio allocations, such as portfolio turnovers and the amount of short-selling. Finally, we examine the economic significance of differences in portfolio volatility by a refined version of the utility-based method introduced in West et al. (1993) and Fleming et al. (2001). This approach provides performance fees (net of transaction costs) that a risk-averse investor would be willing to pay to switch from, for instance, covariance forecasts employing daily returns to HF-based forecasts. To provide finite sample inference for these performance characteristics, we embed the entire evaluation methodology into a stylized "portfolio bootstrap" framework based on a random sampling of asset subsets.

We summarize the major results as follows. First, even naive HF-based forecasts outperform all low-frequency (LF) methods in terms of portfolio volatility. This is particularly true during the turbulent crisis period. Here, an investor with high risk aversion and a daily horizon would be willing to pay up to 199 basis points to benefit from a lower portfolio volatility produced from HF data. This superiority of HF-based forecasts persists up to a monthly horizon with the corresponding performance fee being still 99 basis

points. Second, while eigenvalue cleaning, as applied to BRK estimates by Hautsch et al. (2012), performs well as a robust baseline approach, adaptive or fixed factor structures constitute an effective alternative. Third, short-term smoothing of HF-based covariance matrix estimates can be beneficial for further reducing portfolio volatility. In contrast, smoothing over too long time intervals increases volatility but lowers portfolio turnover. The latter, however, is of importance if the transaction cost level is particularly high. Fourth, constructing forecasts based on a simple dynamic specification of (realized) covariances further improves the performance of HF-based forecasts. During the crisis period, the performance fees an investor with pronounced risk aversion would pay for switching from LF-based predictions amount to 328 and 239 basis points for a daily and monthly horizon, respectively. Fifth, we demonstrate that exploiting HF data for portfolio selection is challenging in a vast investment universe including relatively illiquid assets. In contrast, focusing on the 100 and 30 most heavily-traded stocks out of the S&P 500 universe, we find that basis point fees for switching to HF-based forecasts increase by a multiple.

This chapter contributes to (the few existing) studies on the benefits of HF data for portfolio allocation. In their seminal work, Fleming et al. (2003) apply the evaluation methodology by Fleming et al. (2001) to volatility timing strategies in a general mean-variance context. For a daily forecasting horizon, they find that a risk-averse investor would be willing to pay between 50 and 200 basis points to switch from covariance forecasts based on daily returns to those employing five-minute returns. However, these results are based on allocations across only three highly-liquid futures contracts. Liu (2009) extends the size of the asset universe to 30 by constructing minimum tracking error portfolios (tracking the S&P 500 index) based on the constituents of the Dow Jones Industrial Average. He confirms the benefits of HF-based forecasts in terms of tracking error volatility. Apart from examining the value of HF data for portfolio selection in general, the studies by Bandi et al. (2008) and de Pooter et al. (2008) also aim to determine the optimal intraday sampling frequency. While the former minimize a mean squared error criterion for three S&P 500 stocks and conduct an ex-post economic evaluation, the latter directly compare the performance of volatility timing strategies based on different frequencies considering the constituents of the S&P 100 index.

However, to our best knowledge, no study thoroughly analyzed HF-based forecasts of portfolios covering several hundreds of assets as commonly used in practice. In addition, our contributions to this strand of literature are twofold. First, the above studies are restricted to intraday data sampled at fixed time intervals (e.g., five minutes). We consider the highest frequency possible, employing tailor-made covariance estimators that offer substantial precision gains (see, e.g., Barndorff-Nielsen et al., 2011; Hautsch et al., 2012). Second, the predominant evaluation method is to examine unconditional sample moments of implied portfolio returns (or utilities depending on the latter), which however, can distort the ranking of the underlying covariance matrix forecasts (see Voev, 2009). Our evaluation approach builds upon estimated *conditional* portfolio volatilities, allowing for a more reliable ranking of competing covariance *predictions*.

The remainder of the chapter is organized as follows. Section 3.2 introduces the

general GMV framework, as well as the corresponding evaluation methodology for conditional covariance matrix forecasts. In Section 3.3, we discuss the methods for the construction of conditional covariance predictions based on both HF and LF data. Section 3.4 presents the S&P 500 dataset, more details on the evaluation procedure and the empirical results. Finally, Section 3.5 concludes.

3.2 Global Minimum Variance Portfolios and Covariance Forecasts

The practical implementation of a general mean-variance framework in the spirit of Markowitz (1952) relies on forecasts of the first two conditional moments of asset returns. Consequently, the performance of the predicted (optimal) portfolio allocation depends on the predictability of *both* conditional means and conditional covariances. However, it is well-known that the predictability of first conditional moments of asset returns is much lower than the predictability of conditional (co-)variances (e.g. Merton, 1980). Thus, mean forecasts are subject to substantial prediction errors, which in turn, can completely dominate and distort the analysis (e.g. Michaud, 1989). As a result, isolating the explicit effects of high-dimensional *covariance* forecasts on the resulting portfolio performance is virtually impossible. Hence, in order to eliminate the impact of conditional mean predictions and to solely focus on the value of covariance forecasts, we consider *global* minimum variance portfolios. This proceeding is backed by empirical evidence showing that the noisiness of mean predictions leads to highly unstable portfolio allocations, which are typically outperformed by approaches explicitly avoiding the need of mean forecasts (e.g. DeMiguel et al., 2009; Jagannathan and Ma, 2003; Michaud, 1989). In this sense, our analysis provides insights into the impact of covariance forecasts on portfolio performance without being affected by assumptions or estimation errors associated with mean predictions.

We assume a risk-averse investor with a horizon of h days and an asset universe of m stocks whose optimization problem at day t can be formulated as

$$\min_{w_{t,t+h}} \quad w_{t,t+h}' \, \Sigma_{t,t+h} \, w_{t,t+h} \quad s.t. \quad w_{t,t+h}' \iota = 1, \tag{3.1}$$

where $w_{t,t+h}$ is the $(m \times 1)$ vector of portfolio weights and ι is a $(m \times 1)$ vector of ones. Further, $\Sigma_{t,t+h} := \text{Cov}[r_{t,t+h}|\mathcal{F}_t]$ denotes the $(m \times m)$ conditional covariance matrix of $r_{t,t+h}$, i.e., the $(m \times 1)$ vector of log-returns from day t to $t+h$, given the information set at t, $\mathcal{F}_t$. If, for simplicity, we assume that $\text{Cov}[r_{t+r-1,t+r}, r_{t+s-1,t+s}|\mathcal{F}_t] = 0$, $r, s \geq 1$, $r \neq s$, then $\Sigma_{t,t+h} = \sum_{r=1}^{h} \text{E}[\Sigma_{t+r-1,t+r}|\mathcal{F}_t]$. For $h = 1$, we write $r_{t+1} := r_{t,t+1}$ and, equivalently, $\Sigma_{t+1} := \Sigma_{t,t+1}$. Solving (3.1) yields the GMV portfolio weights given by

$$w_{t,t+h}^{*} = \frac{\Sigma_{t,t+h}^{-1} \, \iota}{\iota' \Sigma_{t,t+h}^{-1} \, \iota}. \tag{3.2}$$

We investigate the benefits of HF data for GMV portfolio selection in terms of forecasts of the conditional covariance matrix, $\widehat{\Sigma}_{t,t+h}$, with corresponding weights $\widehat{w}_{t,t+h}$. To evaluate these predictions, we exploit the basic result of Patton and Sheppard (2008) showing that the conditional variances of the portfolios based on the true conditional covariance matrix $\Sigma_{t,t+h}$ and its forecast $\widehat{\Sigma}_{t,t+h}$ obey

$$\widehat{w}'_{t,t+h}\,\Sigma_{t,t+h}\,\widehat{w}_{t,t+h} > w^{*'}_{t,t+h}\,\Sigma_{t,t+h}\,w^{*}_{t,t+h} \quad \text{if} \quad \widehat{\Sigma}_{t,t+h} \neq \Sigma_{t,t+h}. \tag{3.3}$$

This result yields a natural evaluation criterion, as resulting portfolio variances approach a lower bound if forecasts $\widehat{\Sigma}_{t,t+h}$ approach their population counterparts. Consequently, we consider a forecast $\widehat{\Sigma}_{t,t+h}$ as being "better" if it produces a smaller conditional portfolio variance. As will be discussed below, the conditional portfolio variances can be proxied using HF data.

Importantly, Voev (2009) shows that the above criteria are valid only for conditional, but not unconditional variances. Employing the latter introduces an objective bias, which is driven by the variance of the conditional mean of portfolio returns. Therefore, the bias is negligible only if a mean of zero can be assumed, which is problematic for horizons of more than, e.g., a day. Further, the bias term imposes a penalty on the variation in portfolio weights. This property becomes particularly restrictive when comparing covariance matrix forecasts based on LF and HF data, as intuitively, the latter should be able to incorporate new information faster, implying more variability in the weights. Hence, gains from employing HF data might be understated when unconditional portfolio variances are considered for evaluation.

We assess the economic significance of lower (conditional) portfolio variances by adapting the utility-based evaluation approach suggested by West et al. (1993) and Fleming et al. (2001) to a conditional framework. Accordingly, we assume that the investor has quadratic preferences of the form

$$U\left(r^{p}_{t,t+h}\right) = 1 + r^{p}_{t,t+h} - \frac{\gamma}{2\,(1+\gamma)}\left(1 + r^{p}_{t,t+h}\right)^{2}, \tag{3.4}$$

where $r^{p}_{t,t+h} := \widehat{w}'_{t,t+h}\,r_{t,t+h}$ is the portfolio return, while γ denotes the relative risk aversion. Following Fleming et al. (2003), we consider the two levels $\gamma = 1$ and $\gamma = 10$. For two competing covariance forecasts, $\widehat{\Sigma}^{\mathrm{I}}_{t,t+h}$ and $\widehat{\Sigma}^{\mathrm{II}}_{t,t+h}$, implying the GMV portfolio returns $r^{p,\mathrm{I}}_{t,t+h}$ and $r^{p,\mathrm{II}}_{t,t+h}$, we then determine a value Δ_γ, such that

$$\sum_{t=1}^{T-h} \mathrm{E}\Big[U\Big(r^{p,\mathrm{I}}_{t,t+h}\Big)\Big|\mathcal{F}_t\Big] = \sum_{t=1}^{T-h} \mathrm{E}\Big[U\Big(r^{p,\mathrm{II}}_{t,t+h} - \Delta_\gamma\Big)\Big|\mathcal{F}_t\Big]. \tag{3.5}$$

Δ_γ can be interpreted as a fee the investor would be willing to pay in order to switch from a GMV strategy based on $\widehat{\Sigma}^{\mathrm{I}}_{t,t+h}$ to its counterpart employing $\widehat{\Sigma}^{\mathrm{II}}_{t,t+h}$. As we show in Appendix C.1, the solution to (3.5) depends on the conditional portfolio variances, $\widehat{w}^{i'}_{t,t+h}\,\Sigma_{t,t+h}\,\widehat{w}^{i}_{t,t+h}$, and the conditional means, $\widehat{w}^{i'}_{t,t+h}\,\mu_{t,t+h}$, where $\mu_{t,t+h} := \mathrm{E}[r_{t,t+h}|\mathcal{F}_t]$ is the $(m \times 1)$ vector of conditional expected returns and $i = \mathrm{I}, \mathrm{II}$.

To focus on the effects of differences in (average) conditional portfolio variances, we assume that expected returns are constant over time and identical across all stocks, i.e., $\mu_{t,t+h} = (h/252)\,\mu^{id}\,\iota$, $t = 1, \ldots, T-h$. Then, we obtain the relationship

$$\Delta_\gamma > 0 \quad \text{iff} \quad \overline{\sigma_{\mathrm{I}}^{2,p}} > \overline{\sigma_{\mathrm{II}}^{2,p}}, \qquad \overline{\sigma_i^{2,p}} := \frac{1}{T-h} \sum_{t=1}^{T-h} \widehat{w}_{t,t+h}^{i'}\, \Sigma_{t,t+h}\, \widehat{w}_{t,t+h}^{i}, \quad i = \mathrm{I}, \mathrm{II}, \tag{3.6}$$

under the assumption that $(h/252)\,\mu^{id} \leq 1/\gamma$ (see Appendix C.1).[1] To control for the impact of the assumed level of μ^{id} on the performance fee Δ_γ, we consider a grid of values satisfying the above restriction for the investment horizons and rates of risk aversion employed, i.e., $\mu^{id} \in \{-0.05, 0, 0.05, 0.1\}$. However, as we discuss below, our results are very robust to the specific value of μ^{id}.

3.3 Covariance Estimation and Forecasting in Vast Dimensions

3.3.1 Forecasts Based on High-Frequency Data

Estimating asset return covariances based on high-frequency data requires addressing four major challenges: (i) using high-frequency information based on maximally high sampling frequencies in order to maximize the estimator's efficiency, while (ii) avoiding biases due to microstructure noise (e.g. Hansen and Lunde, 2006) and the asynchronous arrival of observations across assets (e.g. Epps, 1979), as well as (iii) ensuring positive definiteness and (iv) well-conditioning of covariance estimates, i.e., numerical stability of their inverse. Satisfying all criteria simultaneously is challenging, as for instance, fulfilling (i), (iii) and (iv) requires sampling on maximally high frequencies, which in turn, causes substantial biases ruled out by (ii). Conversely, sparse sampling, e.g., based on five-minute returns, as utilized by the classical realized covariance estimator proposed by Andersen et al. (2001), satisfies (ii) but violates (i) and – if the dimension of the portfolio is high – (iv).

A widely-used estimator that is both consistent in the presence of microstructure noise and provides positive semidefinite estimates (thus satisfying (ii) and (iii)) is the multivariate realized kernel proposed by Barndorff-Nielsen et al. (2011). As an important ingredient, this approach involves so-called refresh time sampling for synchronization, which requires to sample prices whenever *all* assets have been traded (i.e., have been refreshed) at least once. This naturally implies a loss of efficiency, since the sampling frequency is driven by those assets trading slowest. As stressed and illustrated by Hautsch et al. (2012), this loss of efficiency can be substantial (thus violating (i)) if the number of

[1] Even in case $(h/252)\,\mu^{id} > 1/\gamma$, we always have that $\Delta_\gamma > 0$ if $\overline{\sigma_{\mathrm{I}}^{2,p}} > \overline{\sigma_{\mathrm{II}}^{2,p}}$. However, the above condition on μ^{id} is not overly restrictive. For the longest investment horizon and highest level of risk aversion we consider, i.e., $h = 20$ and $\gamma = 10$, we need to impose that $\mu^{id} \leq 1.26$. That is, the assumed annualized expected return may not exceed 126 percentage points.

assets and their heterogeneity in terms of trading frequency is high. In the extreme case, covariance matrix estimates might even become ill-conditioned (thus violating (iv)).

The Blocked Realized Kernel

To address this problem and construct estimates which satisfy all criteria, we consider the *blocked* realized kernel put forward by Hautsch et al. (2012). The idea behind the blocked realized kernel is to assign the assets to groups according to their (average) trading frequency and to estimate the underlying correlation matrix groupwise.

In a general framework, we denote the log-price of asset i at time τ by $p_\tau^{(i)}$, $i = 1, \ldots, m$. For the assumptions on the price process that ensure consistency of the (blocked) multivariate realized kernel, we refer to Barndorff-Nielsen et al. (2011). On day t, $t = 1, \ldots, T$, the j-th price observation of asset i is at time $\tau_{t,j}^{(i)}$, where $j = 1, \ldots, N_t^{(i)}$ and $i = 1, \ldots, m$. Let G be the specified number of liquidity groups, yielding the blocks $b = 1, \ldots, B$, with $B = G\left(G+1\right)/2$. Further, we denote the set of indices of the m_b assets associated with block b by $\mathcal{I}_b$. Applying the multivariate realized kernel methodology to the assets in $\mathcal{I}_b$ then requires refresh time sampling with refresh times defined as the time it takes for all the assets in this set to trade or refresh posted prices, i.e.,

$$r\tau_{t,1}^b := \max_{i \in \mathcal{I}_b} \left\{ \tau_{t,1}^{(i)} \right\}, \quad r\tau_{t,l+1}^b := \max_{i \in \mathcal{I}_b} \left\{ \tau_{t,N^{(i)}(r\tau_{t,l}^b)+1}^{(i)} \right\}, \tag{3.7}$$

where $N^{(i)}(\tau)$ denotes the number of price observations of asset i before time τ. Accordingly, vectors of synchronized returns are obtained as

$$r_{t,l}^b := p_{r\tau_{t,l}^b} - p_{r\tau_{t,l-1}^b}, \quad l = 1, \ldots, n_t^b, \tag{3.8}$$

where n_t^b is the number of refresh time observations in block b.

The multivariate realized kernel on block b is defined as

$$K_t^b := \sum_{h=-H_t^b}^{H_t^b} k\left(\frac{h}{H_t^b + 1}\right) \Gamma_t^{h,b}, \tag{3.9}$$

where $k(\cdot)$ is given by the Parzen Kernel and $\Gamma_t^{h,b}$ is an autocovariance matrix, i.e.,

$$\Gamma_t^{h,b} := \begin{cases} \sum_{l=h+1}^{n_t^b} r_{t,l}^b \, r_{t,l-h}^{b'} & \text{for} \quad h \geq 0 \\ \sum_{l=-h+1}^{n_t^b} r_{t,l+h}^b \, r_{t,l}^{b'} & \text{for} \quad h < 0. \end{cases} \tag{3.10}$$

H_t^b is a block-specific smoothing bandwidth that is chosen as in Section 3.4 of Barndorff-Nielsen et al. (2011). Based on (3.9), we compute the corresponding estimate of the correlation block b as

$$R_t^{K,b} := \left(V_t^b\right)^{-1} K_t^b \left(V_t^b\right)^{-1}, \; V_t^b := \operatorname{diag}\left[K_t^{b,(ii)}\right]^{1/2}, \; i = 1, \ldots, m_b, \tag{3.11}$$

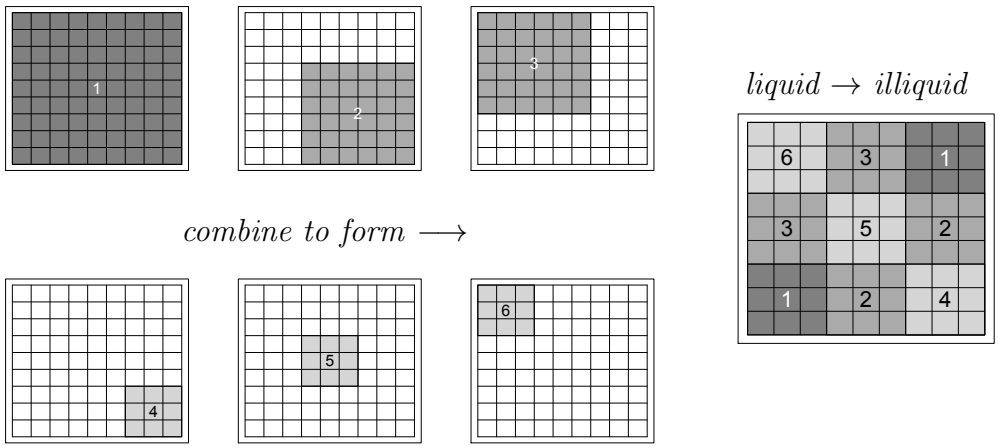


Figure 3.1: Visualization of the Blocking Strategy According to Hautsch et al. (2012)

where $K_t^{b,(ii)},\ i = 1, \ldots, m_b$, are the diagonal elements of K_t^b.

The correlation matrix R_t^{BRK} is then obtained as a hierarchical combination of the correlation blocks $R_t^{K,b},\ b = 1, \ldots, B$. Figure 3.1 from Hautsch et al. (2012) illustrates the blocking strategy in a covariance matrix, where the top-left corner is associated with the most liquid assets and the bottom-right corner is associated with the least liquid assets. The data is decomposed into three equal-sized liquidity groups ($G = 3$), yielding six correlation blocks. Then, in a first step, the entire correlation matrix (block one) is estimated. Subsequently, we obtain estimates of blocks two and three associated with the correlations between the less liquid and more liquid assets, respectively. Finally, blocks four to six contain the correlations *within* each liquidity group. Efficiency gains arise due to a more effective synchronization and thus a higher (refresh time) sampling frequency within each block. Consequently, all blocks – except block one – are estimated with higher precision than in the plain realized kernel. Finally, from the (block-wise) estimated correlation matrix R_t^{BRK}, the BRK estimate of the covariance matrix is constructed according to

$$BRK_t := V_t^{RK}\ R_t^{BRK}\ V_t^{RK},\ V_t^{RK} := \operatorname{diag}\big[RK_t^{(i)}\big]^{1/2},\ i = 1, \ldots, m, \tag{3.12}$$

with $RK_t^{(i)},\ i = 1, \ldots, m$, denoting variance estimates based on the univariate version of the realized kernel (Barndorff-Nielsen et al., 2008a). Consequently, the variance elements are estimated with highest precision, since in a univariate setting synchronization by refresh time sampling is not necessary. We implement the realized kernel estimator following the procedure from Barndorff-Nielsen et al. (2008b).

Smoothing, Regularization and Construction of Forecasts

Variations in portfolio weights require a rebalancing of the portfolio and thus cause transaction costs. The latter can be reduced by keeping covariance matrix forecasts suf-

ficiently stable. The explicit consideration of transaction costs in the underlying portfolio optimization problem, however, results in an empirically challenging problem, as it requires bounding the variability of portfolio weights and thus of the covariance matrix over time. Although the derivation of an explicit solution of this problem is beyond the scope of this chapter, we still aim at studying the impact of competing covariance forecast constructions on the resulting portfolio turnover. A straightforward method to stabilize covariance matrix estimates is to "smooth" them over time by computing simple averages over S days, i.e., $BRK_{t,S} := (1/S) \sum_{s=1}^{S} BRK_{t-s+1}$, where $BRK_{t,1} = BRK_t$.[2] Then, a smoothed correlation matrix is obtained as

$$R_{t,S}^{BRK} := \left(V_{t,S}^{RK}\right)^{-1} BRK_{t,S} \left(V_{t,S}^{RK}\right)^{-1}, \; V_{t,S}^{RK} := \operatorname{diag}\left[RK_{t,S}^{(i)}\right]^{1/2}, \; i = 1, \ldots, m, \quad (3.13)$$

with $RK_{t,S}^{(i)} := (1/S) \sum_{s=1}^{S} RK_{t-s+1}^{(i)}$, $i = 1, \ldots, m$, being smoothed univariate realized kernel estimates.

Estimating correlation matrices block-wise implies efficiency gains, but yields estimates (even after smoothing) which are not guaranteed to be positive semidefinite and well-conditioned. Indefinite matrices feature negative eigenvalues, while ill-conditioned matrices possess eigenvalues that are close to zero, which makes inversions numerically unstable. Particularly for the computation of minimum variance portfolio weights as in (3.2), however, it is crucial that covariance matrices are both positive definite and well-conditioned. These requirements make it necessary to employ suitable regularization techniques.

As a first alternative, we follow Hautsch et al. (2012) employing the eigenvalue cleaning procedure proposed by Laloux et al. (1999). This method rests on the idea of comparing the (empirical) distribution of eigenvalues of the estimated correlation matrix to the theoretical distribution of eigenvalues one would obtain under independence of the m processes. The latter is derived from random matrix theory and yields the expected distribution of eigenvalues assuming these are completely driven by noise. Consequently, eigenvalues strongly departing from the theoretical distribution are identified as "signals" carrying significant information on cross-sectional dependencies. Conversely, eigenvalues being close to zero, and thus below a theoretical upper threshold, are identified as "noisy". They are likely to be noninformative, while causing the correlation matrix to be ill-conditioned. Hence, these eigenvalues can be inflated, making estimates well-conditioned without significantly losing information. See Appendix C.2 for details.

As a second regularization technique, we consider a factor structure based on the spectral components of the correlation matrix. Covariance forecasts based on factor models have been demonstrated to improve the performance of minimum variance portfolios (e.g. Chan et al., 1999). Moreover, a factor structure ensures fast convergence of the factor inverse if the number of factors is small relative to the number of assets (see Fan et al., 2008). Accordingly, we consider a spectral decomposition of the smoothed

[2]Obviously, one might also "smooth" in a more sophisticated way by applying weighting schemes, e.g., based on kernel methods. We leave this for further research but show that even smoothing based on simple averages yields superior results, see Section 3.4.3.

correlation matrix estimate on day t, i.e.,

$$R_{t,S}^{BRK} = Q_{t,S}\,\Lambda_{t,S}\,Q_{t,S}', \tag{3.14}$$

where $\Lambda_{t,S}$ is the diagonal $(m \times m)$ matrix of eigenvalues ordered from largest to smallest, while $Q_{t,S}$ denotes the orthonormal $(m \times m)$ matrix of corresponding eigenvectors. Then, by retaining only the first $k_{t,S} \leq m$ correlation eigenvalues and associated eigenvectors, we obtain the factorized estimate of the correlation matrix

$$R_{t,S,(k_{t,S})}^{BRK} = Q_{t,S,(k_{t,S})}\Lambda_{t,S,(k_{t,S})}Q_{t,S,(k_{t,S})}' + \left(I_m - \mathcal{Q}_{t,S,(k_{t,S})}\right), \tag{3.15}$$

where $\mathcal{Q}_{t,S,(k_{t,S})}$ is a diagonal $(m \times m)$ matrix containing the diagonal elements of $Q_{t,S,(k_{t,S})}\Lambda_{t,S,(k_{t,S})}Q_{t,S,(k_{t,S})}'$. The number of factors $k_{t,S}$ is chosen in two ways. Firstly, we select the number of factors for each day t separately employing the criteria by Bai and Ng (2002). For implementation details, we refer to Appendix C.3. Secondly, we consider a factor structure with the numbers of factors fixed to one or three.

Hence, our combined framework for smoothing and regularizing BRK estimates can be summarized as

$$\widetilde{\Sigma}_{t,S}^{v\mathrm{RnB}} := V_{t,S}^{RK}\; R_{t,S}^{v\mathrm{RnB}}\; V_{t,S}^{RK}, \qquad v \in \{\mathrm{E}, \mathrm{F}, \mathrm{1F}, \mathrm{3F}\}, \tag{3.16}$$

with $R_{t,S}^{v\mathrm{RnB}}$ corresponding to the smoothed correlation matrix estimates from (3.13) regularized by eigenvalue cleaning (E) or by imposing an adaptive (F) or fixed (1F and 3F) factor structure. Following Hautsch et al. (2012), in all cases we regularize only if $R_{t,S}^{BRK}$ is nonpositive definite or ill-conditioned. The latter is defined to be the case if

$$\left|\Lambda_{t,S}^{(1)}/\Lambda_{t,S}^{(m)}\right| > 10 \times m, \tag{3.17}$$

where $\Lambda_{t,S}^{(1)}$ and $\Lambda_{t,S}^{(m)}$ are the largest and smallest eigenvalue of $R_{t,S}^{BRK}$, respectively.

Further possibilities for regularization include, for instance, thresholding techniques (Wang and Zou, 2010). However, the latter rely on a sparsity assumption for the underlying covariance matrix, which is problematic given the strong cross-sectional dependencies typical for equity data. Alternatively, as shown by Jagannathan and Ma (2003), regularization can be achieved by imposing no-short-sale constraints in the portfolio optimization problem (3.1). A related result for general gross portfolio constraints is put forward by Fan et al. (2012b) and applied to evaluate covariance matrix estimates using HF data, e.g., in Fan et al. (2012a). Here, we focus on an unconstrained framework, since it explicitly allows us to compare the performance of different regularization methods and to evaluate the forecasting accuracy not only with respect to the covariance matrix, but also to its *inverse*.

We construct forecasts of $\Sigma_{t,t+h}$ based on the information set $\mathcal{F}_t$ by two alternative approaches. Firstly, we evaluate random walk ("naive") forecasts of the form $\widehat{\Sigma}_{t,t+h} = h\,\widetilde{\Sigma}_{t,S}^{v\mathrm{RnB}}$, which will be referred to as $v\mathrm{RnB}(S)$, $v \in \{\mathrm{E}, \mathrm{F}, \mathrm{1F}, \mathrm{3F}\}$. As an alternative to a pure random walk forecast, we propose a simple dynamic model for unsmoothed covariance matrix estimates. When choosing a suitable dynamic specification

for covariance matrices, positive definiteness of forecasts, model parsimony and ease of implementation are important factors to ensure feasibility in a vast-dimensional setting. To guarantee positive definiteness, we follow Andersen et al. (2003) and Chiriac and Voev (2011) in modeling the Cholesky decomposition of covariance matrix estimates, i.e., $\widetilde{\Sigma}_{t,1}^{v\mathrm{RnB}} = L_t\, L_t'$, where L_t is a lower triangular matrix. As L_t contains $m\,(m+1)\,/2$ distinct elements, we ensure tractability in high dimensions by modeling each row or column of L_t independently. Due to its triangular form, modeling the rows or columns of L_t implies a hierarchical specification of dynamics, depending on the ordering of assets. Consequently, (co-)variances associated with assets being ranked first widely follow their individual dynamics, while volatilities associated with higher ranks are subject to several *joint* dynamics. For instance, in case of row modeling, the volatility of the first asset and, in case of column modeling, all scaled covariances thereof with all other stocks follow independent dynamics.[3] To account for this hierarchy, we order the assets according to their (average) trading frequency during the estimation period.

Let $L_t^{(g\bullet)}$ denote the $(g \times 1)$ vector of elements from the g-th row of L_t and $L_t^{(\bullet g)}$ the $((m-g+1)\times 1)$ vector of elements from the g-th column, $g = 1, \ldots, m$. Dynamic specifications for $L_t^{(g\bullet)}$ and $L_t^{(\bullet g)}$ should capture the well-known persistence properties of volatility processes, which can be achieved by fractionally integrated processes (e.g., Andersen et al., 2003), appropriately mixing different frequencies using, e.g., mixed data sampling (MIDAS) techniques as proposed by Ghysels et al. (2006) or heterogeneous autoregressive (HAR) processes introduced by Corsi (2009). We follow the latter strategy, which is in the spirit of Chiriac and Voev (2011) applying HAR dynamics to the Cholesky factors of realized covariance estimates. Accordingly, we consider the HAR$(1, 5, 20)$ specifications

$$
\begin{aligned}
L_t^{(g\bullet)} &= c^{(g\bullet)} + \alpha_d^{(g\bullet)} L_{t-1}^{(g\bullet)} + \frac{\alpha_w^{(g\bullet)}}{5} \sum_{s=1}^{5} L_{t-s}^{(g\bullet)} + \frac{\alpha_m^{(g\bullet)}}{20} \sum_{s=1}^{20} L_{t-s}^{(g\bullet)} + \varepsilon_t^{(g\bullet)},\ g = 1, \ldots, m, \\
L_t^{(\bullet g)} &= c^{(\bullet g)} + \alpha_d^{(\bullet g)} L_{t-1}^{(\bullet g)} + \frac{\alpha_w^{(\bullet g)}}{5} \sum_{s=1}^{5} L_{t-s}^{(\bullet g)} + \frac{\alpha_m^{(\bullet g)}}{20} \sum_{s=1}^{20} L_{t-s}^{(\bullet g)} + \varepsilon_t^{(\bullet g)},\ g = 1, \ldots, m,
\end{aligned}
\tag{3.18}
$$

where $c^{(g\bullet)}$ and $c^{(\bullet g)}$ are $(g \times 1)$ and $((m-g+1)\times 1)$ parameter vectors, respectively, while the remaining parameters are scalars. We will refer to these specifications as Row- and Column-Cholesky-HAR (RCHAR and CCHAR) models. Based on (least-squares) parameter estimates, the models (3.18) yield h-step ahead forecasts $\widehat{L}_{t+h}^{(g\bullet)}$ and $\widehat{L}_{t+h}^{(\bullet g)}$, $g = 1, \ldots, m$, which are combined to form $\widehat{L}_{t+h}$. Finally, we construct forecasts of $\Sigma_{t,t+h}$ as $\widehat{\Sigma}_{t,t+h} = \sum_{r=1}^{h} \widehat{L}_{t+r}\, \widehat{L}_{t+r}'$. These forecasts involve a bias, as they rely on a nonlinear transformation of the covariance matrix. However, we abstain from a bias correction,

[3] The first row of L_t contains the diagonal element $\sqrt{\widetilde{\Sigma}_{t,1}^{(1,1)}}$, while the first column equals the vector $\left(\widetilde{\Sigma}_{t,1}^{(1,1)}, \widetilde{\Sigma}_{t,1}^{(1,2)}, \ldots, \widetilde{\Sigma}_{t,1}^{(1,m)}\right)' / \sqrt{\widetilde{\Sigma}_{t,1}^{(1,1)}}$.

as e.g., Chiriac and Voev (2011) demonstrate that this bias is empirically negligible. In any case, this issue should be of minor relevance when considering an economic instead of a statistical loss function.

3.3.2 Forecasts Based on Daily Data

We assess the merits of covariance forecasts based on HF data for the portfolio selection framework presented in Section 3.2 by benchmarking the former against methods employing daily returns. A comprehensive overview of these approaches can be found in Sheppard (2012). The three classes of estimators we consider are (i) multivariate GARCH models, (ii) (regularized) rolling window sample covariance matrices and (iii) RiskMetrics. (i) and (ii) have been shown to perform well in the econometric and finance literature, while (iii) is of relevance in financial practice. In this context, we will denote by u_t the $(m \times 1)$ vector of demeaned returns at day t, i.e., $u_t := r_t - \mu_t$, $t = 1, \ldots, T$, where as for the utility-based evaluation above and in line with, e.g., Hansen and Lunde (2005) we assume that the vector of conditional mean returns μ_t is constant over time.

Multivariate GARCH Models

Multivariate GARCH (MGARCH) models parameterize the dynamics of the conditional covariance matrix Σ_{t+1}. For a survey of this model class, we refer to Bauwens et al. (2006). We consider the scalar version of the vector GARCH model (S-VEC) introduced in Bollerslev et al. (1988) and the dynamic conditional correlation (DCC) model proposed by Engle (2002a). The former is motivated by the results on spectral components of covariance and correlation matrices in Zumbach (2009a) that favor a direct modeling of conditional covariance matrices. For that purpose, the S-VEC model is the most parsimonious approach. Employing DCC specifications is justified by their superior out-of-sample prediction accuracy within the MGARCH class when considering various statistical loss functions and different dimensions (e.g. Caporin and McAleer, 2012; Laurent et al., 2012). We estimate both models by Gaussian QML, i.e., assuming $u_{t+1}|\mathcal{F}_t \sim N(0, \Sigma_{t+1})$.

The S-VEC model is a direct extension of the univariate GARCH specification. Ensuring covariance targeting as proposed by Engle and Mezrich (1996), it can be formulated as

$$\Sigma_{t+1} = \bar{\Sigma}\,(1 - \alpha_h - \beta_h) + \alpha_h\, u_t\, u_t' + \beta_h\, \Sigma_t, \quad \alpha_h, \beta_h \geq 0, \quad \alpha_h + \beta_h < 1, \tag{3.19}$$

where $\bar{\Sigma} := \mathrm{E}[u_t\, u_t']$ denotes the unconditional covariance matrix of u_t, which is consistently estimated by the corresponding sample moment. Then, α_h and β_h are estimated by QML using the composite likelihood method proposed by Engle et al. (2008). Accordingly, the joint likelihood is replaced by the sum of pairwise likelihoods, ensuring

tractability in high dimensions.[4] Using the parameter estimates in specification (3.19), we construct h-step ahead forecasts $\widehat{\Sigma}_{t+h}$, yielding $\widehat{\Sigma}_{t,t+h} = \sum_{r=1}^{h} \widehat{\Sigma}_{t+r}$.

The DCC model decomposes the conditional covariance matrix according to

$$\Sigma_{t+1} = V_{t+1}\, R_{t+1}\, V_{t+1}, \quad V_{t+1} := \operatorname{diag}\big[\sigma_{t+1}^{2,(i)}\big]^{1/2}, \quad i = 1, \ldots, m, \tag{3.20}$$

with the conditional variances $\sigma_{t+1}^{2,(i)}$ following univariate GARCH processes, while a similar dynamic structure is imposed on the conditional correlations in R_{t+1}, i.e.,

$$\begin{aligned}
&\sigma_{t+1}^{2,(i)} = \omega_i + \alpha_i\, u_t^{(i),2} + \beta_i\, \sigma_t^{2,(i)}, \quad \omega_i, \alpha_i, \beta_i \geq 0, \quad \alpha_i + \beta_i < 1, \quad i = 1, \ldots, m, \\
&R_{t+1} = \big(V_{t+1}^{z}\big)^{-1}\, Z_{t+1}\, \big(V_{t+1}^{z}\big)^{-1},\ V_{t+1}^{z} := \operatorname{diag}\big[Z_{t+1}^{(ii)}\big]^{1/2}, \quad i = 1, \ldots, m, \\
&Z_{t+1} = \bar{Z}\,(1 - \alpha_z - \beta_z) + \alpha_z\, \epsilon_t\, \epsilon_t' + \beta_z\, Z_t, \quad \alpha_z, \beta_z \geq 0, \quad \alpha_z + \beta_z < 1,
\end{aligned} \tag{3.21}$$

where $Z_{t+1}^{(ii)}$, $i = 1, \ldots, m$, are the diagonal elements of Z_{t+1}, $\epsilon_t := V_t^{-1}\, u_t$ is the $(m \times 1)$ vector of devolatilized returns and $\bar{Z} := \mathrm{E}[\epsilon_t\, \epsilon_t']$. Estimation is carried out in three steps. Firstly, we estimate the m univariate GARCH$(1,1)$ models. Secondly, $\bar{Z}$ is estimated by correlation targeting, i.e., replacing $\mathrm{E}[\epsilon_t\, \epsilon_t']$ with its sample analogue.[5] Finally, we estimate the correlation parameters by the composite likelihood approach. Based on QML parameter estimates and the dynamics in (3.21), one-step ahead covariance forecasts can be straightforwardly constructed as

$$\widehat{\Sigma}_{t+1} = \widehat{V}_{t+1}\, \widehat{R}_{t+1}\, \widehat{V}_{t+1}, \quad \widehat{V}_{t+1} := \operatorname{diag}\big[\widehat{\sigma}_{t+1}^{2,(i)}\big]^{1/2}, \quad i = 1, \ldots, m. \tag{3.22}$$

To obtain the multi-step forecasts necessary for computing $\widehat{\Sigma}_{t,t+h} = \sum_{r=1}^{h} \widehat{\Sigma}_{t+r}$, $h > 1$, we use the approximations suggested in Engle and Sheppard (2005) and Engle (2009, ch. 9.1).

Regularized Rolling Window Sample Covariance

The sample covariance matrix computed from L (demeaned) daily returns is defined as

$$C_t := \frac{1}{L} \sum_{l=1}^{L} u_{t-l+1} u_{t-l+1}'. \tag{3.23}$$

The covariance matrix estimate C_t is positive definite whenever $L \geq m$, but inversion can be numerically unstable even if the latter condition is fulfilled. Accordingly, we regularize C_t using alternative techniques if it is ill-conditioned according to the definition

[4] In our vast-dimensional setting, we follow a suggestion of Engle et al. (2008) and use only adjacent pairs of assets. The results do not change qualitatively when modifying the ordering of assets.

[5] Aielli (2011) shows that the resulting estimator of $\bar{Z}$ is inconsistent and proposes a "corrected" DCC (cDCC) model. However, Caporin and McAleer (2012) find the latter having an inferior forecasting performance compared to the original DCC specification.

in Section 3.3.1. We denote the resulting estimate by C_t^{reg}, where $C_t^{\text{reg}} = C_t$ if no regularization is imposed. Covariance forecasts are then computed as $\widehat{\Sigma}_{t,t+h} = h\, C_t^{\text{reg}}$.

As a simple regularization method, we consider factor models based on the principal components of C_t. The strong performance of factor structures in GMV portfolio applications is documented by Chan et al. (1999), showing that a three-factor model according to the Fama and French (1993) factors is sufficient. While the latter are factors constructed based on asset return characteristics and economic fundamentals, an approximation thereof using principal components can be motivated, for instance, by the results in Connor (1995) on the similar explanatory power of fundamental and statistical factor models. Let $\Lambda^c_{t,(k_t)}$ be the diagonal $(k_t \times k_t)$ matrix of the first k_t eigenvalues and $Q^c_{t,(k_t)}$ the $(m \times k_t)$ matrix of the corresponding eigenvectors of C_t. Then, the resulting factorized covariance matrix estimate is

$$C_t^{\text{reg}} = Q^c_{t,(k_t)} \Lambda^c_{t,(k_t)} Q^{c'}_{t,(k_t)} + \left(V_t^c - \mathcal{Q}^c_{t,(k_t)} \right), \tag{3.24}$$

where V_t^c and $\mathcal{Q}^c_{t,(k_t)}$ are diagonal $(m \times m)$ matrices containing the diagonal elements of C_t and $Q^c_{t,(k_t)} \Lambda^c_{t,(k_t)} Q^{c'}_{t,(k_t)}$, respectively. In the spirit of Chan et al. (1999), we consider a three-factor structure (i.e., $k_t = 3$) and, alternatively, examine a more restrictive framework with $k_t = 1$. Further, we allow for a closer comparison with FRnB estimates by choosing k_t on a dynamic basis using the Bai and Ng (2002) criteria discussed in Appendix C.3.

As a second type of regularization, we use the shrinkage technique initially proposed by Stein (1956) and adopted by Ledoit and Wolf (2003) for sample covariance matrices. The resulting shrunk estimator is a weighted average of C_t and a restricted, positive definite target F_t, i.e.,

$$C_t^{\text{reg}} = \phi\, F_t + (1 - \phi)\, C_t, \quad 0 \leq \phi \leq 1, \tag{3.25}$$

where ϕ is an estimate of the optimal shrinkage intensity derived by Ledoit and Wolf (2003) minimizing the squared error loss. As shrinkage target F_t, they consider the one-factor model by Sharpe (1963), showing that the resulting estimator outperforms, e.g., the pure one-factor and three-factor model. As an approximation, we employ the principal component structure (3.24) with $k_t = 1$. In addition, we follow Ledoit and Wolf (2004) and let F_t be given by the equicorrelation model, i.e., the covariance matrix implied by setting the common correlation equal to the cross-sectional average of all pairwise sample correlations implied by C_t.

Finally, we regularize C_t by the eigenvalue cleaning procedure that is applied to BRK estimates in Section 3.3.1 and discussed in more detail in Appendix C.2. Laloux et al. (2000) demonstrate that sample covariance matrices regularized by this technique yield considerably lower portfolio volatilities than their "uncleaned" counterparts in minimum-variance applications.

RiskMetrics

RiskMetrics covariance forecasts constitute the industry standard. The original RiskMetrics1994 approach is based on an exponentially-weighted moving average (EWMA) of the outer products of demeaned returns, i.e.,

$$\widehat{\Sigma}_{t+1} = \frac{(1-\lambda)}{\left(1-\lambda^{L^{RM}-1}\right)} \sum_{l=1}^{L^{RM}} \lambda^{l-1}\, u_{t-l+1} u'_{t-l+1}, \quad 0 \leq \lambda \leq 1, \tag{3.26}$$

where L^{RM} denotes the window length. We follow the suggestion made in J.P. Morgan/Reuters (1996) for daily returns and set $\lambda = 0.94$. If the forecast $\widehat{\Sigma}_{t+1}$ is ill-conditioned based on the criterion from Section 3.3.1, we apply the tailored regularization technique suggested in Zumbach (2009b), which relies on a two-stage shrinkage. See Zumbach (2009b) for details. Forecasts of $\Sigma_{t,t+h}$ are then computed as $\widehat{\Sigma}_{t,t+h} = h\,\widehat{\Sigma}^{\text{reg}}_{t+1}$, where $\widehat{\Sigma}^{\text{reg}}_{t+1}$ is the regularized forecast with $\widehat{\Sigma}^{\text{reg}}_{t+1} = \widehat{\Sigma}_{t+1}$ if no regularization is necessary.

Additionally, we employ the updated RiskMetrics2006 methodology, which introduces pseudo-long memory dynamics by assuming a hyperbolic decay of the weights on lagged outer products of returns. The corresponding one-step ahead covariance forecast is

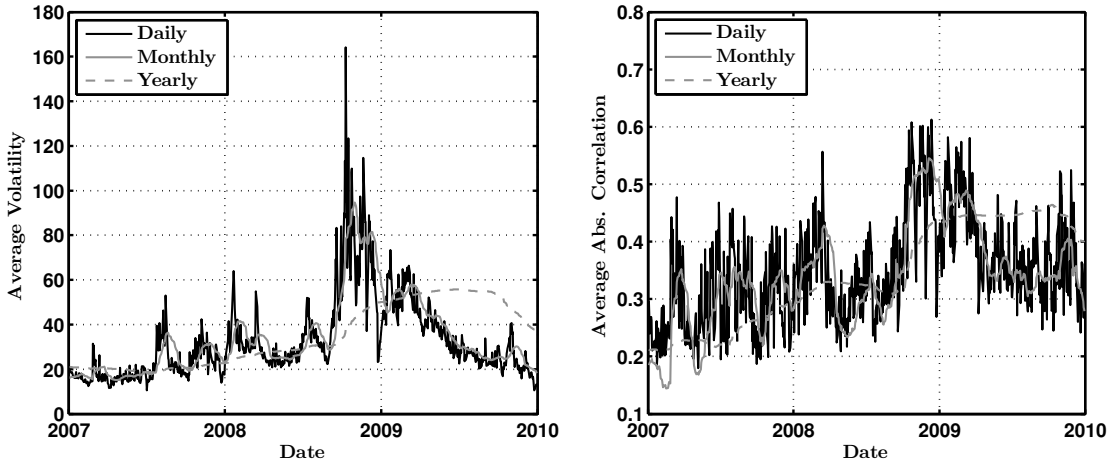
$$\widehat{\Sigma}_{t+1} = \sum_{l=1}^{L^{RM}} \lambda_l\, u_{t-l+1} u'_{t-l+1}, \quad \lambda_l := \sum_{v=1}^{v_{\max}} \zeta_v \frac{(1-\theta_v)}{\left(1-\theta_v^{L^{RM}-1}\right)} \theta_v^{l-1}, \tag{3.27}$$
$$\zeta_v := \frac{1}{D}\left(1 - \frac{\ln(\eta_v)}{\ln(\eta_0)}\right),\ \theta_v := \exp(-1/\eta_v),\ \eta_v := \eta_1\, \rho^{v-1},$$

where the constant D is specified such that $\sum_v \zeta_v = 1$, η_0 is a logarithmic decay factor, while η_1 and $\eta_{v_{\max}}$ denote the lower and upper cut-off, respectively. ρ is an additional tuning parameter and $v_{\max}$ is determined by specifying the values of the other parameters. We use the values suggested in Zumbach (2006), i.e., $\eta_0 = 1560$, $\eta_1 = 4$, $\eta_{v_{\max}} = 512$ and $\rho = \sqrt{2}$. Finally, we construct forecasts of $\Sigma_{t,t+h}$ according to $\widehat{\Sigma}_{t,t+h} = \sum_{r=1}^{h} \widehat{\Sigma}^{reg}_{t+r}$, where multi-step predictions $\widehat{\Sigma}_{t+r}$, $r > 1$, are computed following Appendix A of Zumbach (2006).

3.4 Empirical Results

3.4.1 Data and Empirical Setup

We employ mid-quotes for the constituents of the S&P 500 index extracted from the Trade and Quote (TAQ) database. We focus on the 400 assets with the longest continuous trading history during the sample period between January 2006 and December 2009, covering approximately $1,000$ trading days and including the financial crisis after the bankruptcy filing of Lehman Brothers Inc. We discard the first 15 minutes of each



(a) Average Volatility (b) Average Absolute Correlation

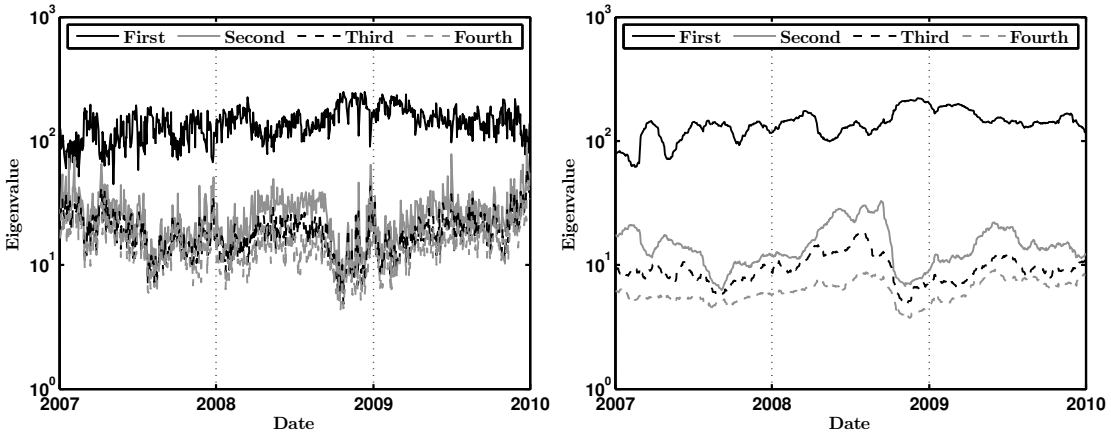
Figure 3.2: Cross-Sectional Averages of Volatility and Absolute Correlation Estimates Based on smoothed BRK estimates using daily, monthly or yearly window. Volatilities are annualized square roots of diagonal elements and are reported in percentage points.

trading day to avoid opening effects and conduct additional steps to clean the raw quote data. Details are provided in Appendix C.4.

Based on the cleaned mid-quote data, we compute BRK estimates as outlined in Section 3.3.1 using $G = 4$ liquidity groups, which will be motivated below. Further, we smooth the BRK estimates over weekly, monthly, quarterly, half-yearly and yearly windows, i.e., $S \in \{1, 5, 20, 63, 126, 252\}$. For three smoothing windows, Figure 3.2 depicts the resulting averages of the square roots of diagonal elements, i.e., volatility estimates, and of the absolute values of pairwise correlations. Two major features are apparent. First, there is a considerable increase of both volatility and absolute correlation during the heyday of the financial crisis in the later part of 2008. Second, employing BRK estimates smoothed over monthly and yearly windows implies a noticeable stabilization. The latter effect can also be confirmed for the eigenvalues of the corresponding correlation matrix estimates displayed in Figure 3.3. Here, smoothing is helpful to separate the dynamics of the first (largest) eigenvalue, which allows for a better signal extraction. The result that the first eigenvalue follows own dynamics different from those of other eigenvalues is at odds with findings based on correlation matrices estimated over long-term rolling windows of *daily* data (e.g. Zumbach, 2009a).

Following Section 3.3.1, we regularize indefinite or ill-conditioned smoothed BRK estimates by eigenvalue cleaning (ERnB) or imposing a factor structure (FRnB, 1FRnB and 3FRnB). As we show in the web appendix in more detail, regularization is necessary for *all* days in the sample and *every* smoothing window.[6] Figure 3.4 gives the number of

[6]The web appendix is available at



(a) Unsmoothed (b) Monthly Smoothing Window

Figure 3.3: Eigenvalues of BRK Correlation Matrix Estimates
Based on logarithmic scale.

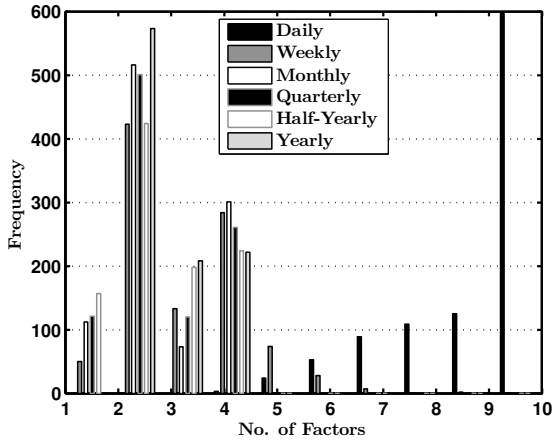


Figure 3.4: Sample Distribution of Number of Factors for FRnB Estimates
Number of factors is determined by applying Bai and Ng (2002) criteria from Appendix C.3 to BRK estimates smoothed over different windows.

factors based on BRK estimates smoothed over different windows resulting from adaptive factor selection using the Bai and Ng (2002) criteria (FRnB). The positive relationship between the length of the smoothing window and the degree of parsimony of the implied factor structure is apparent.

Our analysis focuses on open-to-close covariance matrices, whereby noisy overnight returns do not have to be included. This approach is in line with Andersen et al. (2010b)

`http://amor.cms.hu-berlin.de/~malecpet/MHFDPF_appendix.pdf`.

treating overnight returns as deterministic jumps. Accordingly, we measure the vector of daily returns, r_t, by the vector of open-to-close returns, which can be interpreted as close-to-close returns corrected for the above deterministic jumps. To implement the methods based on daily returns from Section 3.3.2, we obtain the vector of demeaned returns, u_t, by subtracting the sample mean during the respective estimation period.

Using data up to day t, we compute out-of-sample forecasts of the conditional covariance matrices $\Sigma_{t,t+h}$ for daily, weekly and monthly horizons, i.e., $h \in \{1, 5, 20\}$. Rolling window sample covariance matrices are computed using a baseline window length of $L = 252$ days, although alternative window sizes will be examined in a sensitivity analysis. RiskMetrics forecasts are computed employing all available data up to day t with an initial in-sample period of 252 days. Both the sample covariance matrix and RiskMetrics estimates need to be regularized for each day (see web appendix). Finally, we construct covariance forecasts based on MGARCH, as well as R- and CCHAR models using the same expanding windows as for RiskMetrics. R- and CCHAR parameters are re-estimated at each step. In the case of MGARCH models, we estimate the parameters over the entire sample for reasons of numerical stability.

The initial in-sample period comprises the time from 01/2006 to 12/2006. Motivated by the descriptive results above and in order to gain insights into the forecasting performance during "normal" and "non-normal" market periods, we conduct a separate evaluation for a period of 375 days before the financial crisis, covering the time from 01/2007 until 06/2008 ("pre-crisis period"), and the period of 377 days from 07/2008 to 12/2009 including the financial crisis ("crisis period").

3.4.2 Evaluation and Inference in the Portfolio Selection Framework

The forecasts of the conditional covariance matrix, $\widehat{\Sigma}_{t,t+h}$, are used as inputs for the GMV portfolio selection framework in (3.1) and (3.2), yielding the weights $\widehat{w}_{t,t+h}$. The resulting conditional portfolio variance, $\widehat{w}'_{t,t+h}\,\Sigma_{t,t+h}\,\widehat{w}_{t,t+h}$, is then estimated by the five-minute realized portfolio variance

$$\sigma^{2,p}_{t,t+h} := \widehat{w}'_{t,t+h}\,\mathrm{RCov}_{t,t+h}\,\widehat{w}_{t,t+h}, \tag{3.28}$$

where $\mathrm{RCov}_{t,t+h}$ is the five-minute realized covariance from day t to $t+h$, i.e., the sum of outer products of five-minute return vectors obtained by previous-tick interpolation (e.g. Dacorogna et al., 2001). The realized portfolio variances based on competing covariance forecasts are used to compute performance fees Δ_γ, $\gamma \in \{1, 10\}$, according to (3.4) and (3.5).

In addition, we examine several basic characteristics of the GMV portfolio allocations. Following de Pooter et al. (2008), we evaluate portfolio turnover rates to proxy transaction costs proportional to the traded dollar amount for every stock. For a horizon h, the total return of the portfolio from $t-h$ to t is given by $r^p_{t-h,t} := \sum_i \widehat{w}^{(i)}_{t-h,t}\, r^{(i)}_{t-h,t}$, where $\widehat{w}^{(i)}_{t-h,t}$ and $r^{(i)}_{t-h,t}$ are the weight and return of stock i, respectively. Then, before rebalancing to the next period, the weight of stock i in the portfolio changes to

$\widehat{w}_{t-h,t}^{(i)} \left(1 + r_{t-h,t}^{(i)}\right) \left(1 + r_{t-h,t}^{p}\right)^{-1}$. Consequently, the portfolio turnover is given by

$$\mathrm{po}_{t,h} := \sum_{i=1}^{m} \left| \widehat{w}_{t,t+h}^{(i)} - \widehat{w}_{t-h,t}^{(i)} \frac{1 + r_{t-h,t}^{(i)}}{1 + r_{t-h,t}^{p}} \right|. \tag{3.29}$$

Secondly, we quantify the portfolio concentration of resulting GMV portfolio weights. For instance, Oomen (2009) stresses that estimation errors might imply extreme positions and may cause practical pitfalls, such as disproportionate transaction costs or an excessive market impact. We measure portfolio concentration in terms of the norm of the vector of portfolio weights,

$$\mathrm{pc}_{t,h} := \left\| \widehat{w}_{t,t+h} \right\|_2 = \left(\sum_{i=1}^{m} \widehat{w}_{t,t+h}^{(i)\,2} \right)^{1/2}, \tag{3.30}$$

which is minimized for an equally-weighted portfolio, i.e., $\widehat{w}_{t,t+h} = (1/m)\,\iota$. Finally, motivated by the analysis in Liu (2009), we evaluate the size of short positions in the portfolio. Verifying to which extent short-sale constraints would be violated is of practical relevance, since many portfolio managers are prohibited from taking such positions. Hence, we compute the sum of negative portfolio weights as

$$\mathrm{sp}_{t,h} := \sum_{i=1}^{m} \widehat{w}_{t,t+h}^{(i)} \, \mathbb{1}_{\left(\widehat{w}_{t,t+h}^{(i)} < 0\right)}, \tag{3.31}$$

where $\mathbb{1}_{(\cdot)}$ denotes an indicator function.

To assess the statistical significance of performance differences between competing forecasts, we perform a stylized "portfolio bootstrap". Firstly, we create asset indices by drawing random samples of size 350 without replacement from the uniform distribution on the integers $1, \ldots, 400$, which is repeated 1000 times. Second, for each random set of assets and every covariance matrix forecasting model, we compute: (i) the GMV portfolio weights for each horizon and day, (ii) the square root of the sample average of the (annualized) realized portfolio variance in (3.28), $\bar{\sigma}_p^a$, (iii) the resulting annualized performance fees relative to competing forecasts, Δ_γ^a, $\gamma \in \{1, 10\}$, for all considered values of the (identical) conditional mean μ^{id}, as well as (iv) the sample averages of the above portfolio characteristics in (3.29), (3.30), and (3.31), i.e., $\overline{po}$, $\overline{pc}$ and $\overline{sp}$, respectively. For the quantities in (ii)-(iv), we examine median values across all random samples. Additionally, we report the standard deviations of $\bar{\sigma}_p^a$. The empirical implementation of the outlined re-sampling procedure is computationally demanding, as it requires the inversion of more than two million 350×350 covariance matrices for each forecasting method.

3.4.3 The Economic Value of High-Frequency Data

Global Minimum Variance Portfolio Performance

Table 3.1 and 3.2 report the GMV portfolio performance of ERnB, RCHAR and CCHAR forecasts with the latter utilizing unsmoothed ERnB estimates. Throughout the analysis, we fix the number of groups in the blocking strategy to $G = 4$, which will be justified by means of a robustness check below. Table 3.1 and 3.2 also report the performance of factor-based forecasts. For sake of brevity, however, we only show the best-performing factor models minimizing the median realized portfolio volatility for each smoothing window. The complete results are available in the web appendix. The following findings can be summarized.

Firstly, covariance predictions based on a *dynamic* model yield better GMV portfolio performances than those based on a "naive" forecast. Prior to the crisis, the median realized portfolio volatility declines by five standard deviations (s.d.'s) when switching from random walk ERnB(1) to RCHAR forecasts. During the crisis period, the gains induced by dynamic forecasts even increase up to 13 s.d.'s. HAR-based forecasts correspond to weighted averages of past realized covariances and thus are by construction "smoother" in time than random walk forecasts. This property pays off in terms of less volatile portfolio weights and thus lower portfolio turnover. The gains even increase for weekly and monthly forecasts. Further, it turns out that CCHAR forecasts are superior to RCHAR forecasts. In particular, in the pre-crisis period, the difference in median realized portfolio volatility is less than one s.d. for $h = 1$, but during the crisis period, CCHAR forecasts yield a median portfolio volatility that is lower by three s.d.'s. This is also reflected by lower portfolio turnovers induced by CCHAR forecasts.

Secondly, varying the length of the smoothing window has an ambiguous effect. Unsmoothed or only moderately smoothed forecasts result in the lowest portfolio volatility, less short positions and lower portfolio concentration (i.e., more diversification). The benefits of using the most recent data and thus producing forecasts which are highly responsive to new information have to be confronted, however, with a higher variability in portfolio weights, causing a higher portfolio turnover and hence higher transaction costs. These effects yield a natural tradeoff between responsiveness and (too high) variability of covariance forecasts. Not surprisingly, portfolio turnover is minimized by using maximally long smoothing intervals, i.e., one year in our setting.

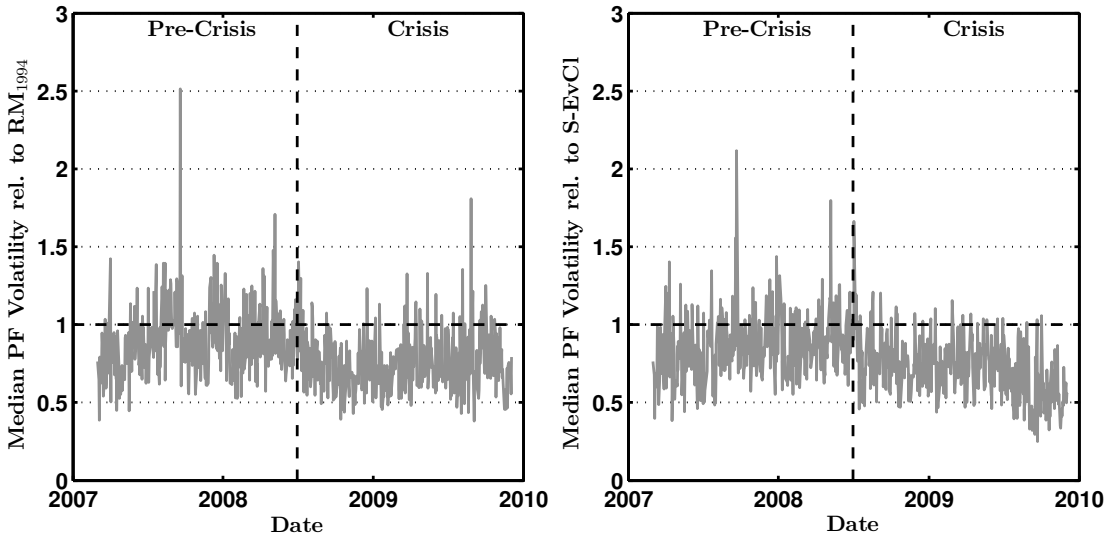
Thirdly, we show that eigenvalue cleaning generally results in the lowest portfolio turnovers and yields less concentrated weights, as well as smaller short positions. Factor-based regularization (FRnB and 3FRnB), however, becomes effective only if the underlying estimates are sufficiently smoothed. In this case, they yield the lowest portfolio volatility and turnover. These effects are particularly apparent during the crisis period. Here, the combination of smoothing and factor-based regularization yields the best portfolio performance in terms of lower portfolio volatility and turnover. In more stable market periods, such as prior to the crisis, the necessity of smoothing and thus the effectiveness of factor-based regularization declines, making eigenvalue cleaning superior. In contrast, factor structures based on unsmoothed BRK estimates result in

Table 3.1: GMV Portfolio Performance of Forecasts Employing High-Frequency Data During Pre-Crisis Period
Medians ($m(\cdot)$) and standard deviations ($s(\cdot)$) across 1,000 random samples of the square root of the annualized average realized portfolio variance ($\bar{\sigma}_p^a$) using predicted GMV weights (in percentage points). Each random sample contains 350 assets out of the entire 400 asset universe. $\overline{\text{po}}$ is the average turnover as defined in (3.29) expressed in percentage points. $\overline{\text{pc}}$ denotes the sample average of the portfolio concentration measure (3.30). $\overline{\text{sp}}$ is the sample average of the sum of negative portfolio weights. RCHAR/CCHAR is based on unsmoothed ERnB estimates. Evaluation is conducted for the pre-crisis period, 01/2007 to 06/2008. Three lowest entries of portfolio volatility and two lowest values of turnover, short position (absolute) and concentration are bold.

	h = 1					**h = 5**					**h = 20**				
	$m(\bar{\sigma}_p^a)$	$s(\bar{\sigma}_p^a)$	$m(\overline{po})$	$m(\overline{sp})$	$m(\overline{pc})$	$m(\bar{\sigma}_p^a)$	$s(\bar{\sigma}_p^a)$	$m(\overline{po})$	$m(\overline{sp})$	$m(\overline{pc})$	$m(\bar{\sigma}_p^a)$	$s(\bar{\sigma}_p^a)$	$m(\overline{po})$	$m(\overline{sp})$	$m(\overline{pc})$
ERnB(1)	**7.49**	0.07	208.77	**-0.60**	**0.21**	**7.88**	0.07	45.30	**-0.60**	**0.21**	8.62	0.01	12.15	**-0.60**	**0.21**
ERnB(5)	7.68	0.08	73.67	**-0.79**	0.23	8.05	0.09	35.97	**-0.79**	0.23	8.53	0.01	11.26	**-0.79**	0.23
ERnB(20)	8.65	0.10	26.75	-0.88	0.24	8.83	0.11	16.37	-0.88	0.24	9.15	0.01	9.44	-0.88	0.24
ERnB(63)	9.54	0.12	11.15	-0.99	0.25	9.64	0.12	7.35	-0.99	0.25	9.95	0.01	4.67	-0.99	0.25
ERnB(126)	10.14	0.12	6.86	-1.03	0.24	10.18	0.12	4.36	-1.03	0.24	10.41	0.01	3.15	-1.03	0.24
ERnB(252)	10.91	0.12	**5.03**	-0.98	0.23	10.98	0.12	**3.04**	-0.98	0.23	11.29	0.01	**2.25**	-0.98	0.23
FRnB(5)	7.62	0.07	84.16	-0.89	0.23	7.98	0.08	39.68	-0.89	0.23	**8.38**	0.01	12.20	-0.89	0.23
3FRnB(20)	8.41	0.10	25.63	-1.02	0.25	8.54	0.10	15.06	-1.02	0.25	8.85	0.01	9.23	-1.02	0.25
3FRnB(63)	9.15	0.11	10.53	-1.13	0.26	9.22	0.11	6.52	-1.13	0.26	9.49	0.01	4.63	-1.13	0.26
3FRnB(126)	9.74	0.12	6.93	-1.18	0.26	9.81	0.12	4.16	-1.18	0.26	10.09	0.01	3.02	-1.18	0.26
3FRnB(252)	10.19	0.13	**5.57**	-1.17	0.26	10.26	0.13	**3.13**	-1.17	0.26	10.60	0.01	**2.23**	-1.17	0.26
RCHAR	**7.14**	0.07	101.69	-0.80	**0.22**	**7.45**	0.07	20.86	-0.80	**0.21**	**8.04**	0.01	5.05	-0.80	**0.21**
CCHAR	**7.18**	0.07	56.80	-0.85	0.23	**7.52**	0.07	18.66	-0.86	0.22	**8.15**	0.01	5.05	-0.86	0.22

Table 3.2: GMV Portfolio Performance of Forecasts Employing High-Frequency Data During Crisis Period
Medians ($m(\cdot)$) and standard deviations ($s(\cdot)$) across 1,000 random samples of the square root of the annualized average realized portfolio variance ($\bar{\sigma}_p^a$) using predicted GMV weights (in percentage points). Each random sample contains 350 assets out of the entire 400 asset universe. $\overline{\text{po}}$ is the average turnover as defined in (3.29) expressed in percentage points. $\overline{\text{pc}}$ denotes the sample average of the portfolio concentration measure (3.30). $\overline{\text{sp}}$ is the sample average of the sum of negative portfolio weights. RCHAR/CCHAR is based on unsmoothed ERnB estimates. Evaluation is conducted for the period including the crisis, 07/2008 to 12/2009. Three lowest entries of portfolio volatility and two lowest values of turnover, short position (absolute) and concentration are bold.

	h = 1					**h = 5**					**h = 20**				
	$m(\bar{\sigma}_p^a)$	$s(\bar{\sigma}_p^a)$	$m(\overline{po})$	$m(\overline{sp})$	$m(\overline{pc})$	$m(\bar{\sigma}_p^a)$	$s(\bar{\sigma}_p^a)$	$m(\overline{po})$	$m(\overline{sp})$	$m(\overline{pc})$	$m(\bar{\sigma}_p^a)$	$s(\bar{\sigma}_p^a)$	$m(\overline{po})$	$m(\overline{sp})$	$m(\overline{pc})$
ERnB(1)	14.02	0.11	215.93	**-0.66**	**0.21**	14.61	0.11	46.69	**-0.66**	**0.21**	15.60	0.01	12.78	**-0.66**	**0.21**
ERnB(5)	13.91	0.11	91.65	**-0.90**	0.25	14.32	0.11	43.41	**-0.90**	0.25	14.98	0.01	12.77	-0.90	0.25
ERnB(20)	14.90	0.14	34.21	-1.06	0.26	15.08	0.14	18.78	-1.06	0.26	15.53	0.01	10.74	-1.06	0.26
ERnB(63)	15.83	0.17	14.29	-1.16	0.28	15.92	0.17	8.76	-1.16	0.28	16.13	0.02	6.11	-1.16	0.28
ERnB(126)	16.32	0.19	12.02	-1.35	0.31	16.40	0.19	7.22	-1.35	0.31	16.54	0.02	4.48	-1.35	0.31
ERnB(252)	16.85	0.20	**9.11**	-1.44	0.32	16.92	0.20	**5.06**	-1.44	0.32	17.02	0.02	**3.42**	-1.44	0.32
FRnB(5)	**13.54**	0.11	111.82	-1.08	0.27	**13.99**	0.11	50.74	-1.08	0.27	**14.68**	0.01	15.06	-1.08	0.27
FRnB(20)	14.46	0.13	36.90	-1.19	0.28	14.63	0.13	20.11	-1.19	0.28	15.13	0.01	11.94	-1.19	0.28
FRnB(63)	15.50	0.15	18.53	-1.31	0.29	15.59	0.15	11.02	-1.31	0.29	15.90	0.01	7.76	-1.31	0.29
FRnB(126)	16.20	0.18	13.90	-1.50	0.32	16.27	0.18	8.12	-1.50	0.32	16.50	0.02	5.27	-1.50	0.32
FRnB(252)	17.13	0.19	**10.13**	-1.59	0.34	17.21	0.19	**5.36**	-1.59	0.34	17.30	0.02	**3.41**	-1.59	0.34
RCHAR	**12.92**	0.12	128.39	-0.89	**0.24**	**13.43**	0.12	28.66	-0.89	**0.24**	**14.38**	0.01	8.19	**-0.88**	**0.24**
CCHAR	**12.55**	0.10	64.77	-0.94	0.25	**12.97**	0.10	25.08	-0.95	0.25	**13.70**	0.01	8.32	-0.95	0.25



(a) Rel. to RM_{1994} (b) Rel. to S-EvCl

Figure 3.5: Median Portfolio Volatility of CCHAR Forecasts Relative to Benchmarks Time series of ratios $m(\sigma_{t,t+1}^{p,\text{CCHAR}})/m(\sigma_{t,t+1}^{p,\text{bench}})$, where $\sigma_{t,t+1}^{p}$ is the square root of the realized portfolio variance in (3.28) computed for $h = 1$. $m(\cdot)$ denotes the median across 1,000 random samples with each random sample containing 350 assets out of the entire 400 asset universe.

highly unstable forecasts and are not reported here (for details, see web appendix).

Table 3.3 and 3.4 show the corresponding results based on forecasting models utilizing *daily* returns as presented in Section 3.3.2. We find that covariance forecasts based on HF data as evaluated in Table 3.1 and 3.2 outperform *all* "low-frequency" (LF) benchmarks up to a weekly horizon. The best-performing LF methods in terms of median portfolio volatility are the RiskMetrics1994 estimator, as well as the rolling window sample covariance matrix regularized by eigenvalue cleaning. The strong performance of the latter, particularly during volatile periods, indicates that the strength of a proper conditioning scheme might be even more important than imposing a dynamic forecasting model. Nevertheless, during the pre-crisis period, (random-walk-type) ERnB(1) forecasts yield a median portfolio volatility which is three s.d.'s lower than the best-performing LF benchmark. This performance gain increases to seven s.d.'s if not naive but (dynamic) RCHAR specifications are used. During the volatile crisis period, the superiority of HF-based approaches becomes even stronger, resulting in a decrease in median realized portfolio volatility of up to 17 s.d.'s in case of a CCHAR model. The dominance of HF-based forecasts, particularly during the crisis period, is graphically highlighted by Figure 3.5, which displays the time series of median portfolio volatility implied by CCHAR forecasts relative to the two best-performing LF benchmarks.

Not surprisingly, the above effects are strongest for daily horizons ($h = 1$) and become

Table 3.3: GMV Portfolio Performance of Forecasts Employing Low-Frequency Data During Pre-Crisis Period
Medians ($m(\cdot)$) and standard deviations ($s(\cdot)$) across 1,000 random samples of the square root of the annualized average realized portfolio variance ($\bar{\sigma}_p^a$) using predicted GMV weights (in percentage points). Each random sample contains 350 assets out of the entire 400 asset universe. $\overline{\text{po}}$ is the average turnover as defined in (3.29) expressed in percentage points. $\overline{\text{pc}}$ denotes the sample average of the portfolio concentration measure (3.30). $\overline{\text{sp}}$ is the sample average of the sum of negative portfolio weights. Forecasts are based on S-VEC and DCC models, regularized RiskMetrics1994 and RiskMetrics2006 estimators (RM_{1994} and RM_{2006}), as well as the rolling window sample covariance matrix of daily returns over 252 days regularized by a one- or three-factor structure (1F or 3F), a factor structure based on the Bai and Ng (2002) criteria (BN-F), eigenvalue cleaning (S-EvCl) and shrinkage towards an equicorrelation or one-factor model (SHRK_{EC} or SHRK_{SF}). In addition, results for the equally-weighted portfolio (EQW) are reported. Evaluation is conducted for the pre-crisis period, 01/2007 to 06/2008. Two lowest entries of portfolio volatility, turnover, short position (absolute) and concentration are bold.

	h = 1					**h = 5**					**h = 20**				
	$m(\bar{\sigma}_p^a)$	$s(\bar{\sigma}_p^a)$	$m(\overline{po})$	$m(\overline{sp})$	$m(\overline{pc})$	$m(\bar{\sigma}_p^a)$	$s(\bar{\sigma}_p^a)$	$m(\overline{po})$	$m(\overline{sp})$	$m(\overline{pc})$	$m(\bar{\sigma}_p^a)$	$s(\bar{\sigma}_p^a)$	$m(\overline{po})$	$m(\overline{sp})$	$m(\overline{pc})$
EQW	15.24	0.09	**0.97**	**0.00**	**0.05**	15.30	0.09	**0.43**	**0.00**	**0.05**	15.58	0.01	**0.23**	**0.00**	**0.05**
1F	9.47	0.09	**6.26**	**-0.65**	**0.18**	9.54	0.09	**3.41**	**-0.65**	**0.18**	9.89	0.01	**2.06**	**-0.65**	**0.18**
3F	8.56	0.08	10.54	-0.83	0.21	8.60	0.08	5.81	-0.83	0.21	8.82	0.01	3.38	-0.83	0.21
BN-F	8.47	0.11	13.77	-1.04	0.24	8.50	0.11	6.81	-1.04	0.24	8.69	0.01	3.55	-1.04	0.24
S-EvCl	**7.82**	0.08	21.35	-1.14	0.26	**7.87**	0.08	10.12	-1.14	0.26	**8.06**	0.01	4.93	-1.14	0.26
SHRK_{EC}	8.38	0.08	27.74	-1.67	0.33	8.42	0.08	14.16	-1.67	0.33	8.63	0.01	7.55	-1.67	0.33
SHRK_{SF}	7.92	0.08	24.89	-1.44	0.29	7.98	0.08	12.55	-1.44	0.29	8.19	0.01	6.48	-1.44	0.29
RM_{1994}	**7.71**	0.08	34.96	-1.18	0.27	**7.80**	0.08	18.07	-1.18	0.27	**8.04**	0.01	9.40	-1.18	0.27
RM_{2006}	7.98	0.08	38.94	-1.67	0.33	8.08	0.08	18.93	-1.69	0.33	8.40	0.01	9.17	-1.75	0.34
S-VEC	10.11	0.16	37.20	-3.11	0.54	10.12	0.16	19.49	-3.10	0.53	10.24	0.02	10.50	-3.08	0.53
DCC	9.40	0.10	124.13	-2.12	0.41	9.43	0.10	53.69	-2.15	0.41	9.58	0.01	19.52	-2.21	0.41

Table 3.4: GMV Portfolio Performance of Forecasts Employing Low-Frequency Data During Crisis Period
Medians ($m(\cdot)$) and standard deviations ($s(\cdot)$) across 1,000 random samples of the square root of the annualized average realized portfolio variance ($\bar{\sigma}_p^a$) using predicted GMV weights (in percentage points). Each random sample contains 350 assets out of the entire 400 asset universe. $\overline{\text{po}}$ is the average turnover as defined in (3.29) expressed in percentage points. $\overline{\text{pc}}$ denotes the sample average of the portfolio concentration measure (3.30). $\overline{\text{sp}}$ is the sample average of the sum of negative portfolio weights. Forecasts are based on S-VEC and DCC models, regularized RiskMetrics1994 and RiskMetrics2006 estimators (RM_{1994} and RM_{2006}), as well as the rolling window sample covariance matrix of daily returns over 252 days regularized by a one- or three-factor structure (1F or 3F), a factor structure based on the Bai and Ng (2002) criteria (BN-F), eigenvalue cleaning (S-EvCl) and shrinkage towards an equicorrelation or one-factor model (SHRK_{EC} or SHRK_{SF}). In addition, results for the equally-weighted portfolio (EQW) are reported. Evaluation is conducted for the period including the crisis, 07/2008 to 12/2009. Two lowest entries of portfolio volatility, turnover, short position (absolute) and concentration are bold.

	h = 1					**h = 5**					**h = 20**				
	$m(\bar{\sigma}_p^a)$	$s(\bar{\sigma}_p^a)$	$m(\overline{po})$	$m(\overline{sp})$	$m(\overline{pc})$	$m(\bar{\sigma}_p^a)$	$s(\bar{\sigma}_p^a)$	$m(\overline{po})$	$m(\overline{sp})$	$m(\overline{pc})$	$m(\bar{\sigma}_p^a)$	$s(\bar{\sigma}_p^a)$	$m(\overline{po})$	$m(\overline{sp})$	$m(\overline{pc})$
EQW	31.90	0.20	**1.61**	**0.00**	**0.05**	31.88	0.20	**0.71**	**0.00**	**0.05**	31.73	0.02	**0.38**	**0.00**	**0.05**
1F	17.92	0.18	**5.98**	**-0.58**	**0.17**	18.07	0.18	**2.92**	**-0.58**	**0.17**	18.41	0.02	**1.73**	**-0.58**	**0.17**
3F	16.89	0.16	13.59	-0.92	0.22	17.13	0.16	6.79	-0.92	0.22	17.44	0.02	3.93	-0.92	0.22
BN-F	15.43	0.13	20.38	-1.33	0.28	15.58	0.13	9.87	-1.33	0.28	**15.85**	0.01	4.92	-1.33	0.28
S-EvCl	**14.93**	0.14	26.39	-1.46	0.32	**15.08**	0.14	12.59	-1.46	0.32	**15.34**	0.01	6.06	-1.46	0.32
SHRK_{EC}	15.53	0.15	32.40	-1.89	0.37	15.70	0.15	16.19	-1.89	0.37	15.98	0.02	8.28	-1.89	0.37
SHRK_{SF}	15.59	0.15	35.98	-1.99	0.37	15.77	0.15	17.79	-1.99	0.37	16.07	0.02	9.07	-1.99	0.37
RM_{1994}	**15.20**	0.13	42.30	-1.42	0.31	**15.44**	0.13	22.12	-1.42	0.31	15.92	0.01	11.61	-1.42	0.31
RM_{2006}	15.36	0.14	48.31	-2.11	0.40	15.58	0.14	23.94	-2.13	0.40	15.97	0.01	11.57	-2.18	0.40
S-VEC	18.47	0.27	40.77	-3.35	0.57	18.45	0.27	20.65	-3.34	0.56	18.30	0.03	10.79	-3.30	0.56
DCC	17.73	0.18	126.41	-1.92	0.41	17.85	0.17	54.35	-1.93	0.41	18.10	0.02	20.32	-1.95	0.40

weaker for longer forecasting horizons. However, although the informational advantage of HF data naturally declines with the prediction interval, we still identify performance gains from HF data even at a monthly horizon. While in the pre-crisis period, the best LF and HF one-month forecast yield exactly the same median portfolio volatility, the latter can be significantly reduced during the crisis if HF-based forecasts are employed.

The dominance of HF-based approaches is obviously due to the efficient use of more recent information, making forecasts more responsive and adaptable to structural changes. These effects particularly pay off during highly volatile periods, such as in 2008. Moreover, we show that HF-based forecasts also yield less concentrated (and thus more diversified) positions, as well as less short-selling. However, as stressed above, the downside of a higher responsiveness of forecasts is a higher variability in portfolio weights, which increases portfolio turnover and transaction costs. The latter could be reduced at the expense of a higher portfolio volatility by using longer, i.e., at least quarterly, smoothing windows. Addressing this tradeoff more thoroughly is a challenging avenue for further research, but is clearly beyond the scope of the current study.

Finally, we also evaluate the performance of a naive investment strategy assigning equal weights ($1/m$) to all assets. Interestingly, the $1/m$-portfolio yields a significantly higher median volatility than all other methods. This finding is at odds with the study of DeMiguel et al. (2009) reporting that strategies based on covariance matrix forecasts cannot consistently outperform a naive diversification strategy. However, it has to be noted that DeMiguel et al. examine unconditional Sharpe ratios, while our evaluation focuses on the *conditional* portfolio volatility (approximated by the *realized* volatility).

Economic Significance

We evaluate the economic gains of employing HF-based covariance forecasts using the utility-based evaluation approach in (3.4) and (3.5). To incorporate the effect of transaction costs, we follow de Pooter et al. (2008) assuming that the latter are proportional to portfolio turnover. Accordingly, (3.5) is extended by defining performance fees net of the difference in transaction costs between the two competing strategies, i.e., $\Delta_\gamma^c := \Delta_\gamma - c\left(\overline{\text{po}}^{\text{II}} - \overline{\text{po}}^{\text{I}}\right)$, where c denotes the proportional transaction costs on each traded dollar and $\overline{\text{po}}^i$ is the (average) turnover implied by the GMV strategy based on the covariance forecasts $\widehat{\Sigma}_{t,t+h}^i$, $i = \text{I}, \text{II}$. However, to avoid assumptions on the level c, we focus on "break-even" trading cost levels implying $\Delta_\gamma^c = 0$ and thus $c_\gamma^* := \Delta_\gamma / \left(\overline{\text{po}}^{\text{II}} - \overline{\text{po}}^{\text{I}}\right)$. The economic interpretation depends on the signs of the performance fee Δ_γ and the turnover difference $\mathcal{D}_{\text{po}} := \overline{\text{po}}^{\text{II}} - \overline{\text{po}}^{\text{I}}$. If $\Delta_\gamma > 0$, $\mathcal{D}_{\text{po}} > 0$ implies that c_γ^* yields the *maximal* level of *positive* transaction costs under which the risk-averse investor is still willing to pay for employing strategy II instead of I, while for $\mathcal{D}_{\text{po}} < 0$, c_γ^* gives the *minimal* level (in absolute terms) of *negative* transaction costs, i.e., transaction credits, under which this is no longer the case. In contrast, given that $\Delta_\gamma < 0$, c_γ^* denotes the *minimal positive* (for $\mathcal{D}_{\text{po}} < 0$) or *negative* (for $\mathcal{D}_{\text{po}} > 0$) transaction cost level necessary to make strategy II superior to strategy I.

Table 3.5 and 3.6 report the median values of the (annualized) performance fees Δ_γ^a in

basis points (bp) the investor would pay in order to switch from the best LF benchmarks to HF-based forecasting methods. Moreover, we show the median values of the corresponding annualized break-even transaction costs c^*_γ. The underlying expected returns are assumed to be identical across stocks and are fixed to $\mu^{id} = 0.05$ (annualized). In the web appendix, we demonstrate that alternative values of μ^{id} yield quantitatively almost identical results. As LF benchmarks, we choose those strategies minimizing the median portfolio volatility or turnover. Among HF-based forecasts, for each smoothing window we select the regularization method yielding the lowest median portfolio volatility. The corresponding findings for all other models are given in the web appendix.

The major observations are as follows. Firstly, by utilizing HF-based covariance forecasts, a risk-averse investor can achieve noticeable economic gains which become substantial during the crisis period. Before the crisis and for a daily horizon, an investor with low (high) risk aversion would be willing to pay 2 (17) bp to switch from the best LF strategy to the best random-walk-type HF forecast (ERnB(1)) and 4 (40) bp to switch to a CCHAR forecast. During the crisis period, these values increase to 20 (199) bp in the naive (FRnB(5)) and 33 (328) bp in the dynamic case. Focusing on longer forecasting intervals, these gains become smaller, however are still substantial even for a monthly horizon if the investor exhibits a high risk aversion. In the latter case, the median performance fees for switching to FRnB(5) and CCHAR forecasts amount to 99 and 238 bp, respectively. Figure 3.6 shows the nonparametrically estimated performance fee densities resulting from the underlying portfolio bootstrap approach. The plots confirm the statistical significance of the results, particularly during the crisis period. Moreover, CCHAR covariance forecasts yield slightly less dispersed performance fee distributions than random-walk-type FRnB(5) forecasts.

Secondly, using HF data remains valuable for more risk averse investors even in the presence of transaction costs. During the crisis period, the annualized median break-even transaction costs associated with the above performance fees for the daily horizon are 0.2 (2) percentage points (pp) for FRnB(5) and 0.9 (9) pp for CCHAR forecasts in case of low (high) risk aversion. These are the median values of the transaction cost levels at which the net performance fee paid by a risk-averse investor for switching from the low-volatility LF benchmark to the HF-based forecasts would just remain positive. When benchmarking against the LF-based forecast yielding the lowest turnover, i.e., the rolling window sample covariance regularized by a one-factor structure, the median break-even transaction costs associated with the CCHAR specification increase to 1.4 (14) pp, which is moderate compared to the increase in the corresponding performance fees. This finding is obviously induced by the low portfolio turnover implied by the one-factor structure, naturally decreasing the impact of transaction costs.

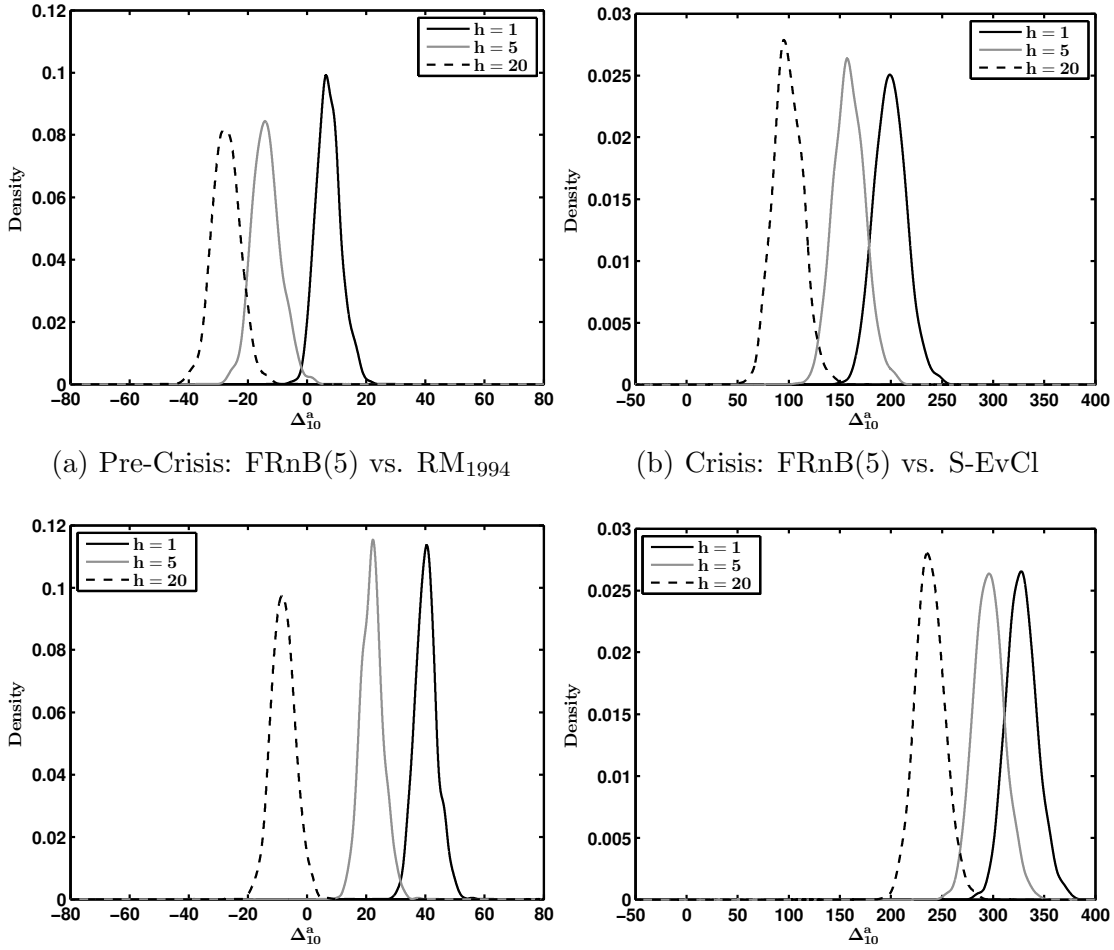
Finally, in several cases, we observe a combination of negative (median) performance fees and positive (median) break-even transaction costs. Here, the explicit consideration of transaction costs favors HF-based covariance forecasts if these costs *exceed* a certain level. For instance, ERnB(252) forecasts yield negative median performance fees vis-a-vis the low-volatility LF benchmark regardless of the level of risk aversion. However, after the introduction of transaction costs of at least 1.8 pp in case of low risk aversion

Table 3.5: Basis Point Fees for Switching from Low-Frequency to High-Frequency Forecasts During Pre-Crisis Period
Medians ($m(\cdot)$) across 1,000 random samples of annualized basis point fees (Δ_γ^a) a risk-averse investor with quadratic utility and relative risk aversion γ would pay to switch from covariance forecasts using daily data to high-frequency-based forecasts. We assume that the constant conditional mean return is identical across all stocks and set it to $\mu^{id} = 0.05$ (annualized). Also reported are break-even transaction costs (c_γ^*) in percentage points, defined as the ratio of Δ_γ^a and the difference of average portfolio turnovers. Each random sample contains 350 assets out of the entire 400 asset universe. Evaluation is conducted for the pre-crisis period, 01/2007 to 06/2008. The low-frequency benchmarks are the RiskMetrics1994 estimator (RM_{1994}) and the sample covariance computed over 252 days regularized by imposing a one-factor model (1F). Two highest entries of the performance fee and the corresponding break-even transaction costs are bold.

	h = 1				**h = 5**				**h = 20**			
	$m(\Delta_1^a)$	$m(c_1^*)$	$m(\Delta_{10}^a)$	$m(c_{10}^*)$	$m(\Delta_1^a)$	$m(c_1^*)$	$m(\Delta_{10}^a)$	$m(c_{10}^*)$	$m(\Delta_1^a)$	$m(c_1^*)$	$m(\Delta_{10}^a)$	$m(c_{10}^*)$
						vs. RM_{1994}						
ERnB(1)	**1.68**	**0.01**	**16.85**	**0.10**	**-0.64**	**-0.00**	**-6.43**	**-0.04**	-4.85	-0.09	-48.57	-0.91
FRnB(5)	0.70	0.01	7.06	0.10	-1.40	-0.01	-14.00	-0.15	-2.73	-0.05	-27.36	-0.53
3FRnB(20)	-5.66	0.69	-56.74	6.86	-6.05	0.45	-60.58	4.53	-6.84	1.69	-68.53	16.95
3FRnB(63)	-12.17	0.54	-121.94	5.41	-12.08	0.23	-121.06	2.27	-12.72	0.14	-127.42	1.43
3FRnB(126)	-17.67	0.67	-177.10	6.75	-17.60	0.27	-176.40	2.70	-18.49	0.15	-185.33	1.55
3FRnB(252)	-22.18	0.80	-222.28	8.00	-22.17	0.31	-222.18	3.14	-23.80	0.18	-238.50	1.76
CCHAR	**4.02**	**0.18**	**40.22**	**1.79**	**2.19**	**0.72**	**21.95**	**7.23**	**-0.81**	**0.01**	**-8.14**	**0.11**
						vs. 1F						
ERnB(1)	**16.77**	**0.08**	**167.91**	**0.83**	**14.43**	**0.07**	**144.56**	**0.69**	11.71	0.06	117.33	0.59
FRnB(5)	15.78	0.20	158.05	2.01	13.69	0.07	137.10	0.74	13.80	0.07	138.22	0.69
3FRnB(20)	9.43	0.45	94.42	4.53	9.06	0.14	90.71	1.42	9.69	0.07	97.09	0.65
3FRnB(63)	2.94	0.46	29.46	4.62	3.01	0.12	30.12	1.24	3.81	0.06	38.13	0.61
3FRnB(126)	-2.59	-5.71	-25.97	-57.25	-2.54	-1.01	-25.44	-10.16	-1.99	-0.15	-19.92	-1.49
3FRnB(252)	-7.09	12.01	-71.08	120.38	-7.09	5.96	-71.04	59.76	-7.27	-2.52	-72.82	-25.26
CCHAR	**19.09**	**0.38**	**191.12**	**3.77**	**17.26**	**0.22**	**172.81**	**2.25**	**15.71**	**0.26**	**157.36**	**2.65**

Table 3.6: Basis Point Fees for Switching from Low-Frequency to High-Frequency Forecasts During Crisis Period
Medians ($m(\cdot)$) across 1,000 random samples of annualized basis point fees (Δ_γ^a) a risk-averse investor with quadratic utility and relative risk aversion γ would pay to switch from covariance forecasts using daily data to high-frequency-based forecasts. We assume that the constant conditional mean return is identical across all stocks and set it to $\mu^{id} = 0.05$ (annualized). Also reported are break-even transaction costs (c_γ^*) in percentage points, defined as the ratio of Δ_γ^a and the difference of average portfolio turnovers. Each random sample contains 350 assets out of the entire 400 asset universe. Evaluation is conducted for the period including the crisis, 07/2008 to 12/2009. The low-frequency benchmarks are given by the sample covariance computed over 252 days regularized by eigenvalue cleaning (S-EvCl) and by imposing a one-factor model (1F). Two highest entries of the performance fee and the corresponding break-even transaction costs are bold.

	h = 1				**h = 5**				**h = 20**			
	$m(\Delta_1^a)$	$m(c_1^*)$	$m(\Delta_{10}^a)$	$m(c_{10}^*)$	$m(\Delta_1^a)$	$m(c_1^*)$	$m(\Delta_{10}^a)$	$m(c_{10}^*)$	$m(\Delta_1^a)$	$m(c_1^*)$	$m(\Delta_{10}^a)$	$m(c_{10}^*)$
						vs. S-EvCl						
ERnB(1)	13.20	0.07	132.25	0.70	6.94	0.04	69.51	0.41	-4.12	-0.03	-41.30	-0.31
FRnB(5)	**19.91**	**0.23**	**199.39**	**2.33**	**15.85**	**0.08**	**158.70**	**0.83**	**9.84**	**0.05**	**98.54**	**0.55**
FRnB(20)	6.88	0.66	68.90	6.59	6.68	0.18	66.95	1.78	3.04	0.03	30.47	0.26
FRnB(63)	-8.70	1.11	-87.21	11.15	-7.94	1.00	-79.51	10.04	-8.90	-0.26	-89.14	-2.61
FRnB(126)	-19.77	1.59	-198.11	15.89	-18.73	0.83	-187.67	8.36	-18.57	1.17	-186.07	11.75
ERnB(252)	-30.62	1.78	-306.96	17.80	-29.55	0.79	-296.15	7.89	-27.29	0.52	-273.52	5.16
CCHAR	**32.74**	**0.86**	**327.73**	**8.60**	**29.55**	**0.47**	**295.85**	**4.75**	**23.74**	**0.53**	**237.69**	**5.29**
						vs. 1F						
ERnB(1)	62.46	0.30	624.99	2.98	56.72	0.26	567.59	2.60	48.10	0.22	481.45	2.18
FRnB(5)	**69.13**	**0.65**	**691.55**	**6.53**	**65.65**	**0.27**	**656.84**	**2.75**	**62.01**	**0.23**	**620.43**	**2.33**
FRnB(20)	56.08	1.81	561.14	18.15	56.37	0.66	564.03	6.57	55.20	0.27	552.40	2.70
FRnB(63)	40.49	3.23	405.26	32.34	41.90	1.04	419.43	10.37	43.23	0.36	432.72	3.58
FRnB(126)	29.44	3.72	294.74	37.28	31.05	1.20	310.84	11.98	33.51	0.47	335.46	4.74
ERnB(252)	18.65	5.96	186.78	59.64	20.25	1.89	202.74	18.91	24.86	0.74	248.91	7.37
CCHAR	**81.90**	**1.40**	**819.15**	**13.98**	**79.38**	**0.72**	**793.98**	**7.18**	**75.93**	**0.58**	**759.54**	**5.77**



(a) Pre-Crisis: FRnB(5) vs. RM_{1994}

(b) Crisis: FRnB(5) vs. S-EvCl

(c) Pre-Crisis: CCHAR vs. RM_{1994}

(d) Crisis: CCHAR vs. S-EvCl

Figure 3.6: Kernel Estimates of Performance Fee Density
Kernel density estimates across 1,000 random samples of the annualized basis point fee (Δ^a_γ) a risk-averse investor with quadratic utility and relative risk aversion $\gamma = 10$ would pay to switch from covariance forecasts using daily data to high-frequency-based forecasts. Each random sample contains 350 assets out of the entire 400 asset universe. The assumed constant conditional mean return is identical across all stocks and set to $\mu^{id} = 0.05$ (annualized). Density estimates are based on the Gaussian kernel and the rule-of-thumb bandwidth with normal reference.

Table 3.7: No. of Liquidity Groups and GMV Portfolio Volatility of ERnB(1) Forecasts Medians ($m(\cdot)$) and standard deviations ($s(\cdot)$) across 1,000 random samples of the square root of the annualized average realized portfolio variance ($\bar{\sigma}_p^a$) using predicted GMV weights for the horizon $h = 1$ (in percentage points). Each random sample contains 350 assets out of the entire 400 asset universe. G denotes the number of liquidity groups used to compute BRK estimates. Evaluation is conducted for the pre-crisis period, 01/2007 to 06/2008, and the period including the crisis, 07/2008 to 12/2009.

	Pre-Crisis		**Crisis**	
G	$m(\bar{\sigma}_p^a)$	$s(\bar{\sigma}_p^a)$	$m(\bar{\sigma}_p^a)$	$s(\bar{\sigma}_p^a)$
1	8.38	0.28	14.43	0.11
2	8.25	0.29	14.25	0.11
4	7.49	0.07	14.02	0.11
5	8.15	0.30	13.98	0.11
8	8.13	0.30	13.94	0.11
10	8.12	0.30	13.93	0.11

and 18 pp in case of high risk aversion, the net performance fee turns positive. These effects materialize whenever the smoothing window is sufficiently long, driving down the turnover of HF-based approaches compared to their LF competitors.

3.4.4 Sensitivity Analysis and Robustness Checks

Number of Liquidity Groups

In the above analysis, we employed $G = 4$ liquidity groups to compute BRK estimates, which in turn, were used to construct HF-based covariance matrix forecasts. Thus, in order to justify this choice, we examine the optimal value of G in terms of median realized portfolio volatility.

Focusing on a daily horizon, Table 3.7 reports the forecasting performance of unsmoothed BRK estimates regularized by eigenvalue cleaning (ERnB(1)) for different values of G. Before the crisis, using four liquidity groups ($G = 4$) yields the lowest volatility. During the crisis period, median portfolio volatility declines monotonously when increasing G. However, for more than four liquidity groups, the magnitude of the reductions exhibits a noticeable decay, as the latter do not even amount to one standard deviation. These results are in line with Hautsch et al. (2012) who find that blocking-based efficiency gains are mainly due to the (more general) separation between liquid and illiquid assets, implying that a relatively low number of liquidity groups is sufficient.

Length of the Estimation Window

The results in Section 3.4.3 suggest a positive relationship between the length of the smoothing window and the resulting portfolio volatility in the case of HF-based covariance matrix forecasts. For the regularized rolling window sample covariance matrix of daily returns, an estimation window of 252 days was employed. Hence as a robustness check, we consider the alternative window lengths of 378, 126, 63 and 20 days and investigate the impact on the median performance fees for switching to HF-based predictions, as well as on the corresponding median break-even transaction costs. We focus on FRnB(5) and ERnB(252) forecasts, representing slight and heavy smoothing, respectively. For each window length of the sample covariance matrix, we then consider both "low-volatility" and "low-turnover" LF-based predictions by choosing the regularization methods that imply the lowest median portfolio volatility or turnover.

Table 3.8 and 3.9 report the results focusing on the crisis period. The corresponding analysis for the pre-crisis sample along with the complete results of the above benchmark selection procedure can be found in the web appendix. For the low-volatility benchmarks, reducing the window length from 252 to, ultimately, 20 days implies a severe precision loss, as the median performance fees for switching to both FRnB(5) and ERnB(252) forecasts increase sharply. In line with intuition, the implied portfolio turnover rises considerably, which can be detected by the disproportionate increase in the median break-even transaction costs. An extension of the estimation window to 378 days causes only a small reduction of the median performance fees, indicating only mild precision gains due to the larger sample size.

In case of low-turnover benchmarks, which employ the one-factor structure for regularization, a decreasing window length is associated with shrinking median performance fees for switching to forecasts using HF data. This finding suggests that, if a particularly restrictive regularization method is considered, the smaller number of observations used for covariance estimation is outweighed by the fact that only the most recent and hence relevant information is utilized. However, the median performance fees remain considerably higher than in case of low-volatility benchmarks based on the "optimal" window lengths. Interestingly, shortening the estimation window does not necessarily imply an excessive rise in portfolio turnover, as the median break-even transaction costs vis-a-vis FRnB(5) forecasts decrease. However, when compared to the more severely smoothed ERnB(252) forecasts, the growing turnover becomes apparent. As long as the performance fee is positive, median break-even transaction costs increase or even become negative, which corresponds to a situation where LF-based covariance forecasts cause a higher (average) turnover than their HF counterparts. Equivalently for negative performance fees, (positive) median break-even transaction costs contract.

Table 3.8: Impact of Estimation Window on Basis Point Fees for Switching from Low-Frequency to FRnB(5) Forecasts
Medians ($m(\cdot)$) across 1,000 random samples of the annualized basis point fee (Δ_γ^a) a risk-averse investor with quadratic utility and relative risk aversion γ would pay to switch from covariance forecasts using regularized sample covariances of daily data computed over different windows to high-frequency-based forecasts. We assume that the constant conditional mean return is identical across all stocks and set it to $\mu^{id} = 0.05$ (annualized). Also reported are break-even transaction costs (c_γ^*) in percentage points, defined as the ratio of Δ_γ^a and the difference of average portfolio turnovers. Each random sample contains 350 assets out of the entire 400 asset universe. Evaluation is conducted for the period including the crisis, 07/2008 to 12/2009. For each window length of the sample covariance, we consider the regularization yielding the lowest median realized portfolio volatility ("low-volatility") or median portfolio turnover ("low-turnover"). Low-volatility benchmarks: eigenvalue cleaning (378, 252 and 126 days), shrinkage towards single-factor model (63) and shrinkage towards equicorrelation model (20). Low-turnover benchmarks: imposing one-factor model (all windows). Entries corresponding to window of 252 days and window yielding smallest performance fee are bold.

	h = 1				**h = 5**				**h = 20**			
Wind.	$m(\Delta_1^a)$	$m(c_1^*)$	$m(\Delta_{10}^a)$	$m(c_{10}^*)$	$m(\Delta_1^a)$	$m(c_1^*)$	$m(\Delta_{10}^a)$	$m(c_{10}^*)$	$m(\Delta_1^a)$	$m(c_1^*)$	$m(\Delta_{10}^a)$	$m(c_{10}^*)$
vs. Low-Volatility Benchmarks												
378	19.91	0.21	199.40	2.15	15.65	0.08	156.71	0.75	9.71	0.05	97.27	0.47
252	**19.91**	**0.23**	**199.39**	**2.33**	**15.85**	**0.08**	**158.70**	**0.83**	**9.84**	**0.05**	**98.54**	**0.55**
126	22.95	0.32	229.77	3.22	19.04	0.12	190.63	1.17	14.04	0.11	140.63	1.11
63	26.52	0.39	265.54	3.88	23.99	0.16	240.24	1.65	22.12	0.28	221.55	2.82
20	31.98	0.59	320.17	5.89	29.55	0.26	295.82	2.63	28.01	0.76	280.42	7.59
vs. Low-Turnover Benchmarks												
378	71.72	0.67	717.44	6.75	67.81	0.28	678.44	2.82	62.30	0.23	623.37	2.30
252	**69.13**	**0.65**	**691.55**	**6.53**	**65.65**	**0.27**	**656.84**	**2.75**	**62.01**	**0.23**	**620.43**	**2.33**
126	61.72	0.59	617.59	5.93	60.18	0.26	602.12	2.58	62.34	0.25	623.78	2.47
63	48.12	0.48	481.64	4.77	47.19	0.21	472.29	2.10	52.59	0.22	526.27	2.24
20	**40.39**	**0.47**	**404.35**	**4.73**	**37.58**	**0.20**	**376.17**	**2.00**	**38.84**	**0.23**	**388.82**	**2.32**

Table 3.9: Impact of Estimation Window on Basis Point Fees for Switching from Low-Frequency to ERnB(252) Forecasts Medians ($m(\cdot)$) across 1,000 random samples of the annualized basis point fee (Δ_γ^a) a risk-averse investor with quadratic utility and relative risk aversion γ would pay to switch from covariance forecasts using regularized sample covariances of daily data computed over different windows to high-frequency-based forecasts. We assume that the constant conditional mean return is identical across all stocks and set it to $\mu^{id} = 0.05$ (annualized). Also reported are break-even transaction costs (c_γ^*) in percentage points, defined as the ratio of Δ_γ^a and the difference of average portfolio turnovers. Each random sample contains 350 assets out of the entire 400 asset universe. Evaluation is conducted for the period including the crisis, 07/2008 to 12/2009. For each window length of the sample covariance, we consider the regularization yielding the lowest median realized portfolio volatility ("low-volatility") or median portfolio turnover ("low-turnover"). Low-volatility benchmarks: eigenvalue cleaning (378, 252 and 126 days), shrinkage towards single-factor model (63) and shrinkage towards equicorrelation model (20). Low-turnover benchmarks: imposing one-factor model (all windows). Entries corresponding to window of 252 days and window yielding smallest performance fee are bold.

	h = 1				**h = 5**				**h = 20**			
Wind.	$m(\Delta_1^a)$	$m(c_1^*)$	$m(\Delta_{10}^a)$	$m(c_{10}^*)$	$m(\Delta_1^a)$	$m(c_1^*)$	$m(\Delta_{10}^a)$	$m(c_{10}^*)$	$m(\Delta_1^a)$	$m(c_1^*)$	$m(\Delta_{10}^a)$	$m(c_{10}^*)$
vs. Low-Volatility Benchmarks												
378	**-30.66**	**2.98**	**-307.32**	**29.90**	**-29.85**	**1.43**	**-299.22**	**14.30**	**-27.45**	**1.02**	**-275.13**	**10.20**
252	**-30.62**	**1.78**	**-306.96**	**17.80**	**-29.55**	**0.79**	**-296.15**	**7.89**	**-27.29**	**0.52**	**-273.52**	**5.16**
126	-27.70	0.88	-277.67	8.86	-26.54	0.40	-266.00	4.04	-23.28	0.22	-233.30	2.21
63	-24.01	0.70	-240.65	7.04	-21.55	0.26	-215.98	2.61	-15.15	0.10	-151.77	0.98
20	-18.59	0.38	-186.27	3.85	-15.85	0.14	-158.87	1.38	-9.19	0.05	-92.04	0.47
vs. Low-Turnover Benchmarks												
378	21.32	5.80	213.46	58.08	22.37	1.80	224.02	18.07	25.11	0.67	251.38	6.67
252	**18.65**	**5.96**	**186.78**	**59.64**	**20.25**	**1.89**	**202.74**	**18.91**	**24.86**	**0.74**	**248.91**	**7.37**
126	11.19	8.09	112.10	81.07	14.73	2.72	147.56	27.24	25.17	1.25	252.07	12.50
63	-2.37	1.35	-23.78	13.49	1.71	-0.57	17.12	-5.67	15.48	6.45	155.01	64.59
20	**-10.13**	**0.58**	**-101.53**	**5.81**	**-7.87**	**0.19**	**-78.87**	**1.91**	**1.74**	**-0.03**	**17.42**	**-0.27**

Dimension of the Asset Universe

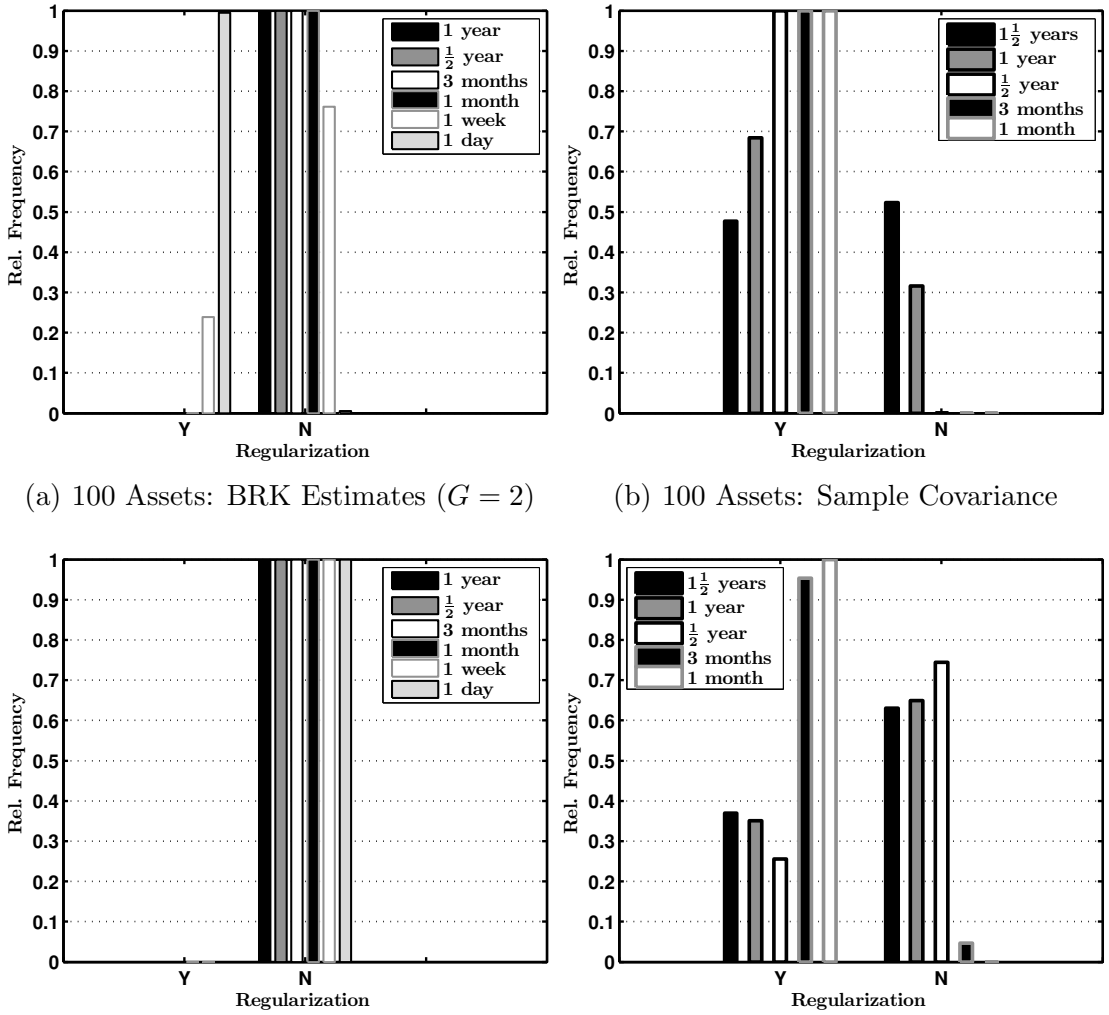
Above, we consider a high-dimensional asset universe comprising 400 stocks, which exhibit a considerable heterogeneity with respect to their liquidity.[7] In order to examine the extent to which the gains from HF data depend on both the problem dimension and the liquidity of the underlying assets, we repeat the analysis for subsets containing the 100 or 30 stocks with the highest number of mid-quote revisions during the sample period. The chosen cross-sectional dimensions correspond to those of the S&P 100 and the Dow Jones Industrial Average, which, e.g., constitute the asset universes for the studies by de Pooter et al. (2008) and Liu (2009), respectively.

For the portfolio bootstrap procedure outlined in Section 3.4.2, we draw random samples containing asset indices of size 85 or 25. The covariance matrix forecasting approaches from Section 3.3.1 and 3.3.2 are implemented as above with three exceptions. First, we compute BRK estimates employing a smaller number of liquidity groups G, i.e., $G = 2$ and $G = 1$ in the 100 and 30 asset case, respectively. As is shown in the web appendix, up to these values of G reductions in median realized portfolio volatility amounting to at least one standard deviation can be achieved. Second, the parameters of MGARCH models are estimated on a day-by-day basis using expanding estimation windows as for R- and CCHAR specifications above. In the 30 asset case, we also consider the full quasi-likelihood instead of the composite likelihood approach. Finally, we account for the fact that regularization of BRK estimates, as well as of rolling window sample covariance and RiskMetrics forecasts is not always imposed according to the conditions discussed in Section 3.3.1 and 3.3.2. For BRK estimates and the LF sample covariance, Figure 3.7 shows that the regularization frequencies are positively related to the dimension and negatively related to the length of the smoothing or estimation window.[8] Thus, we additionally compute forecasts based on unconditional regularization, i.e., independent from the above rule.

The results of the entire analysis can be found in the web appendix. Here, we focus on the median performance fees for switching from the best LF forecasts to random-walk-type HF-based predictions during the crisis period, which are reported in Table 3.10 and 3.11. We choose the two types of best-performing LF benchmarks, as well as the optimal regularization (conditional or unconditional) of HF forecasts for each smoothing window as before. The implied low-volatility benchmarks in the 100 and 30 asset case are given by the sample covariance shrunk unconditionally towards an equicorrelation model and the DCC specification, respectively. The latter fact indicates that MGARCH models are more suitable for moderate dimensions than for vast-dimensional settings. Regarding HF predictions, an unconditional regularization is not advantageous for any smoothing window in both dimensions. Accordingly, no regularization is imposed for smoothing windows of one month or more in the 100 asset case and for all window lengths when

[7] The average number of mid-quote revisions in the cleaned dataset is about $5,000$ for the most liquid stocks and only 250 for the least liquid assets.

[8] The relative regularization frequency of RiskMetrics2006 forecasts drops to around 50% only in the 30 asset case. RiskMetrics1994 forecasts are always regularized (see web appendix).



(a) 100 Assets: BRK Estimates ($G = 2$)

(b) 100 Assets: Sample Covariance

(c) 30 Assets: BRK Estimates ($G = 1$)

(d) 30 Assets: Sample Covariance

Figure 3.7: Regularization Frequency in 100 and 30 Asset Universe
Depending on length of smoothing (BRK) or estimation window (rolling window sample covariance). (Smoothed) BRK estimates are regularized if any correlation eigenvalue is negative or the condition number of the correlation matrix is greater than 10×100 or 10×30. The rolling window sample covariance of daily returns is regularized if the condition number of the corresponding correlation matrix is greater than the above thresholds.

30 assets are considered (cf. Figure 3.7). We refer to the corresponding unregularized random-walk-type forecasts based on a S-day smoothing window as BRK(S).

The first major result is that, in general, the median basis point fees vis-a-vis the low-volatility benchmarks increase considerably. In case of 100 assets and a daily horizon, the median performance fees for switching to ERnB(5) forecasts assuming low (high) risk aversion are 57 (567) bp, which is almost three times the corresponding values for FRnB(5) predictions in the vast-dimensional setting. Very similar results are obtained for 30 assets. The increased benefits from HF data can be explained by the fact that we focus on more liquid assets featuring a higher number of mid-quote revisions, which translates into more precise BRK estimates. The latter effect also implies that, secondly, we observe large median performance fees even at a monthly horizon. Given a high risk aversion, the median fees for switching to BRK(20) forecasts are 427 and 526 bp in the 100 and 30 asset scenario, respectively, which is more than four and five times the highest fee for this horizon found in Section 3.4.3.

Thirdly, the median basis point fees remain positive when employing the three longest smoothing windows regardless of the magnitude of risk aversion or the investment horizon. This finding is of practical importance, as the corresponding forecasts yield a relatively low portfolio turnover, resulting in negative median break-even transaction costs. The latter implies that a risk-averse investor is willing to pay for switching to long-term smoothed HF-based forecasts given any positive transaction cost level. In addition, the fact that, compared to the vast-dimensional scenario, the reduction in median performance fees is less pronounced when moving from short to yearly smoothing windows indicates a higher persistence of the conditional covariance matrix process in the lower dimensional case.

Finally, we do not observe the same gains due to the reduced cross-sectional dimension and even detect some losses when examining performance fees vis-a-vis the low-turnover benchmark with the latter being given by the one-factor model, i.e., the sample covariance matrix unconditionally regularized by a one-factor structure. This result might be explained by the less restrictive nature of the one-factor model if only 100 or 30 assets are considered. However, the same instance implies that the resulting portfolio turnover increases relatively to HF-based forecasts employing longer smoothing windows (see web appendix). For the 30 asset setting, in particular, the latter effect is evidenced by the median break-even transaction costs becoming considerably negative.

3.5 Conclusion

This chapter provides insights into the value of high-frequency (HF) data for short-horizon large-scale portfolio allocation decisions. For that purpose, we construct global minimum variance (GMV) portfolios from the constituents of the S&P 500 index with weights being determined by different conditional covariance matrix forecasts. We consider HF-based forecasts originating from covariance estimates based on the blocked realized kernel proposed by Hautsch et al. (2012). The estimates are smoothed, regularized by either eigenvalue cleaning or imposing a factor structure and, finally, used to

Table 3.10: Basis Point Fees for Switching from Low-Frequency to High-Frequency Forecasts in 100 Asset Universe
Medians ($m(\cdot)$) across 1,000 random samples of the annualized basis point fee (Δ_γ^a) a risk-averse investor with quadratic utility and relative risk aversion γ would pay to switch from covariance forecasts using daily data to high-frequency-based forecasts. We assume that the constant conditional mean return is identical across all stocks and set it to $\mu^{id} = 0.05$ (annualized). Also reported are break-even transaction costs (c_γ^*) in percentage points, defined as the ratio of Δ_γ^a and the difference of average portfolio turnovers. Each random sample contains 85 assets out of the 100 asset universe. Evaluation is conducted for the period including the crisis, 07/2008 to 12/2009. The low-frequency benchmarks are given by the sample covariance matrix computed over 252 days and unconditionally regularized by shrinkage towards an equicorrelation model (SHRK_{EC}^*) or by imposing a one-factor model (1F^*). Two highest entries of the performance fee and the corresponding break-even transaction costs are bold.

	h = 1				**h = 5**				**h = 20**			
	$m(\Delta_1^a)$	$m(c_1^*)$	$m(\Delta_{10}^a)$	$m(c_{10}^*)$	$m(\Delta_1^a)$	$m(c_1^*)$	$m(\Delta_{10}^a)$	$m(c_{10}^*)$	$m(\Delta_1^a)$	$m(c_1^*)$	$m(\Delta_{10}^a)$	$m(c_{10}^*)$
						vs. SHRK_{EC}^*						
3FRnB(1)	40.33	0.28	403.75	2.79	33.24	0.24	332.79	2.39	16.85	0.14	168.76	1.38
ERnB(5)	**56.60**	**0.89**	**566.36**	**8.89**	**49.37**	**0.36**	**494.09**	**3.63**	36.49	0.27	365.34	2.68
BRK(20)	**56.99**	**8.37**	**570.29**	**83.80**	**53.37**	**2.26**	**534.14**	**22.58**	**42.68**	**0.61**	**427.22**	**6.14**
BRK(63)	45.84	-6.49	458.77	-64.94	43.76	-3.33	437.97	-33.30	**36.88**	**-4.94**	**369.14**	**-49.50**
BRK(126)	38.63	-4.05	386.67	-40.59	37.08	-1.79	371.22	-17.92	31.46	-1.13	314.93	-11.27
BRK(252)	35.13	-3.08	351.72	-30.86	34.34	-1.24	343.79	-12.43	31.26	-0.62	312.92	-6.24
						vs. 1F^*						
3FRnB(1)	44.08	0.28	441.21	2.84	36.58	0.22	366.15	2.19	19.21	0.11	192.37	1.10
ERnB(5)	**60.10**	**0.81**	**601.36**	**8.10**	**52.64**	**0.32**	**526.84**	**3.21**	38.80	0.21	388.35	2.05
BRK(20)	**60.54**	**3.51**	**605.78**	**35.11**	**56.76**	**1.12**	**567.94**	**11.16**	**44.94**	**0.37**	**449.80**	**3.69**
BRK(63)	49.60	14.64	496.43	146.47	47.19	3.37	472.30	33.77	**39.24**	**0.87**	**392.79**	**8.75**
BRK(126)	42.43	46.68	424.69	467.27	40.51	6.38	405.50	63.88	33.89	1.38	339.26	13.86
BRK(252)	38.80	-40.53	388.40	-405.68	37.76	-58.41	377.96	-584.79	33.74	14.17	337.79	141.83

Table 3.11: Basis Point Fees for Switching from Low-Frequency to High-Frequency Forecasts in 30 Asset Universe
Medians ($m(\cdot)$) across 1,000 random samples of the annualized basis point fee (Δ_γ^a) a risk-averse investor with quadratic utility and relative risk aversion γ would pay to switch from covariance forecasts using daily data to high-frequency-based forecasts. We assume that the constant conditional mean return is identical across all stocks and set it to $\mu^{id} = 0.05$ (annualized). Also reported are break-even transaction costs (c_γ^*) in percentage points, defined as the ratio of Δ_γ^a and the difference of average portfolio turnovers. Each random sample contains 25 assets out of the 30 asset universe. Evaluation is conducted for the period including the crisis, 07/2008 to 12/2009. The low-frequency benchmarks are the DCC model, as well as the sample covariance matrix computed over 252 days and unconditionally regularized by imposing a one-factor model (1F*). Two highest entries of the performance fee and the corresponding break-even transaction costs are bold.

	h = 1				**h = 5**				**h = 20**			
	$m(\Delta_1^a)$	$m(c_1^*)$	$m(\Delta_{10}^a)$	$m(c_{10}^*)$	$m(\Delta_1^a)$	$m(c_1^*)$	$m(\Delta_{10}^a)$	$m(c_{10}^*)$	$m(\Delta_1^a)$	$m(c_1^*)$	$m(\Delta_{10}^a)$	$m(c_{10}^*)$
						vs. DCC						
BRK(1)	46.13	0.50	461.71	4.96	37.69	0.66	377.26	6.66	30.74	0.85	307.81	8.51
BRK(5)	**56.27**	**-7.36**	**563.08**	**-73.62**	**53.32**	**-9.04**	**533.55**	**-90.43**	**52.54**	**-3.65**	**525.81**	**-36.47**
BRK(20)	**48.11**	**-1.68**	**481.54**	**-16.86**	**50.38**	**-0.89**	**504.21**	**-8.88**	**52.55**	**-1.17**	**525.86**	**-11.67**
BRK(63)	34.22	-1.01	342.54	-10.13	38.45	-0.52	384.86	-5.23	**45.10**	**-0.54**	**451.37**	**-5.37**
BRK(126)	22.11	-0.64	221.39	-6.37	26.81	-0.35	268.47	-3.47	34.20	-0.36	342.34	-3.57
BRK(252)	14.91	-0.42	149.32	-4.17	20.61	-0.26	206.40	-2.55	30.25	-0.28	302.89	-2.82
						vs. 1F*						
BRK(1)	62.33	0.49	623.60	4.95	47.98	0.36	480.23	3.64	25.04	0.19	250.72	1.88
BRK(5)	**72.64**	**2.84**	**726.69**	**28.39**	**63.68**	**0.89**	**637.12**	**8.94**	**46.89**	**0.56**	**469.29**	**5.60**
BRK(20)	**64.38**	**15.35**	**644.12**	**153.59**	**60.78**	**3.38**	**608.18**	**33.78**	**46.97**	**0.90**	**470.15**	**8.98**
BRK(63)	50.39	-44.33	504.29	-443.62	48.56	34.44	486.03	344.63	39.52	2.98	395.59	29.81
BRK(126)	38.24	-17.69	382.75	-177.07	36.89	-14.90	369.34	-149.13	28.55	11.07	285.81	110.80
BRK(252)	31.19	-10.78	312.29	-107.97	30.95	-5.31	309.88	-53.19	25.28	-2.46	253.12	-24.64

construct both random-walk-type predictions and forecasts relying on a simple autoregressive specification. We employ an extensive set of benchmark approaches based on daily returns and compare the competing forecasting methods in terms of estimated conditional portfolio volatility and additional portfolio characteristics. We allow for basic inference by using a "portfolio bootstrap" procedure and investigate the economic gains of reduced portfolio volatility by means of a conditional version of the methodology put forward in West et al. (1993) and Fleming et al. (2001).

Based on mid-quote data from 2006 to 2009, we show the following major results. First, HF-based covariance forecasts outperform low-frequency (LF) approaches over investment horizons of up to a month. The gains in terms of reduced portfolio volatility are considerably larger during the volatile market period including the 2008 financial crisis and are of substantial economic value from the point of view of an investor with pronounced risk aversion. Second, short-term smoothing can be beneficial in terms of lower portfolio volatility, while long-term smoothing always helps to reduce transaction costs. Third, the performance of HF-based strategies can be further improved if naive random-walk-type forecasts are replaced by predictions relying on (even simple) dynamic models. Finally, we find that incorporating methods using HF data into a large-scale portfolio allocation framework is a relatively demanding task. If we focus on subsets comprising only the most liquid S&P 500 assets, the gains from GMV strategies employing HF-based covariance forecasts increase to a considerable extent.

Possible avenues for future research are threefold. First, alternative regularization methods could be considered. Recent examples are the subsampled principal component approach put forward by Abadir et al. (2012) or nonlinear shrinkage as proposed in Ledoit and Wolf (2012). Second, while our choice of a dynamic model for HF-based covariance matrix estimates was mainly driven by parsimony and ease of estimation, richer specifications could be employed. In this context, utilizing HF data in a GARCH framework, as e.g., suggested by Hansen et al. (2010) and Noureldin et al. (2012), appears particularly promising. Further possibilities, also specifically for vast-dimensional settings, are presented in Andersen et al. (2011). Finally, the naive smoothing of covariance matrix estimates could be replaced by an optimal smoothing scheme that strikes a balance between the accuracy of forecasts, implying low portfolio volatility, and the minimization of transaction costs caused by variation in portfolio weights. For this purpose, the approach recently proposed by Kirby and Ostdiek (2012) could be adapted to a HF framework.

Bibliography

Abadir, K. M., W. Distaso, and F. Zikes (2012): "Design-Free Estimation of Large Variance Matrices," Working paper, Imperial College London.

Aielli, G. P. (2011): "Dynamic Conditional Correlation: On Properties and Estimation," "Marco Fanno" Working Papers 0142, Dipartimento di Scienze Economiche "Marco Fanno".

Allen, D., F. Chan, M. McAleer, and S. Peiris (2008): "Finite Sample Properties of the QMLE for the Log-ACD Model: Application to Australian Stocks," *Journal of Econometrics*, 147, 163–185.

Andersen, T. G., T. Bollerslev, P. F. Christoffersen, and F. X. Diebold (2011): "Financial Risk Measurement for Financial Risk Management," in *Handbook of the Economics of Finance Vol. II*, ed. by G. Constanides, M. Harris, and R. Stulz, Amsterdam: Elsevier, forthcoming.

Andersen, T. G., T. Bollerslev, and F. X. Diebold (2010a): "Parametric and Nonparametric Measurements of Volatility," in *Handbook of Financial Econometrics*, ed. by Y. Ait-Sahalia and L. Hansen, Amsterdam: North Holland, 67–137.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001): "The Distribution of Realized Exchange Rate Volatility," *Journal of the American Statistical Association*, 96, 42–55.

——— (2003): "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579–625.

Andersen, T. G., T. Bollerslev, P. Frederiksen, and M. O. Nielsen (2010b): "Continuous-Time Models, Realized Volatilities, and Testable Distributional Implications for Daily Stock Returns," *Journal of Applied Econometrics*, 25, 233–261.

Bai, J. and S. Ng (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221.

Bandi, F., J. R. Russell, and Y. Zhu (2008): "Using High-Frequency Data in Dynamic Portfolio Choice," *Econometric Reviews*, 27, 163–198.

Barndorff-Nielsen, O., P. Hansen, A. Lunde, and N. Shephard (2008a): "Designing Realized Kernels to Measure the Ex-Post Variation of Equity Prices in the Presence of Noise," *Econometrica*, 76, 1481–1536.

——— (2008b): "Realised Kernels in Practice: Trades and Quotes," *Econometrics Journal*, 4, 1–32.

——— (2011): "Multivariate Realised Kernels: Consistent Positive Semi-Definite Estimators of the Covariation of Equity Prices with Noise and Non-Synchronous Trading," *Journal of Econometrics*, 162, 149–169.

Barndorff-Nielsen, O. and N. Shephard (2002): "Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models," *Journal of the Royal Statistical Society, Ser. B.*, 64, 253–280.

——— (2004): "Econometric Analysis of Realized Covariation: High Frequency Based Covariance, Regression, and Correlation in Financial Economics," *Econometrica*, 72, 885–925.

Bauwens, L., F. Galli, and P. Giot (2003): "The Moments of Log-ACD Models," CORE Discussion Paper.

Bauwens, L. and P. Giot (2000): "The Logarithmic ACD Model: An Application to the Bid-Ask Quote Process of Three NYSE Stocks," *Annales D'Economie et de Statistique*, 60, 117–149.

Bauwens, L., P. Giot, J. Grammig, and D. Veredas (2004): "A Comparison of Financial Duration Models via Density Forecasts," *International Journal of Forecasting*, 20, 589–609.

Bauwens, L. and N. Hautsch (2008): "Modeling Financial High-Frequency Data with Point Processes," in *Handbook of Financial Time Series*, ed. by T. G. Andersen, R. A. Davis, J.-P. Kreiss, and T. Mikosch, Heidelberg: Springer, 23–68.

Bauwens, L., S. Laurent, and J. V. K. Rombouts (2006): "Multivariate GARCH Models: A Survey," *Journal of Applied Econometrics*, 21, 79–109.

Berkowitz, J. (2001): "Testing Density Forecasts, with Applications to Risk Management," *Journal of Business & Economic Statistics*, 19, 465–474.

Bollerslev, T., R. F. Engle, and J. M. Wooldridge (1988): "A Capital Asset Pricing Model with Time-Varying Covariances," *Journal of Political Economy*, 96, 116–131.

Bouezmarni, T. and J. V. K. Rombouts (2010): "Nonparametric Density Estimation for Positive Time Series," *Computational Statistics & Data Analysis*, 54, 245–261.

Bouezmarni, T. and O. Scaillet (2005): "Consistency of Asymmetric Kernel Density Estimators and Smoothed Histograms with Application to Income Data," *Econometric Theory*, 21, 390–412.

Bowman, K. O. and L. R. Shenton (1975): "Omnibus Test Contours for Departures from Normality Based on $\sqrt{b_1}$ and b_2," *Biometrika*, 62, 243–250.

Brockwell, A. (2007): "Universal Residuals: A Multivariate Transformation," *Stat Probab Lett.*, 77, 1473–1478.

Brownlees, C. T., F. Cipollini, and G. M. Gallo (2010): "Intra-Daily Volume Modeling and Prediction for Algorithmic Trading," *Journal of Financial Econometrics*, 8, 1–30.

Caporin, M. and M. McAleer (2012): "Robust Ranking of Multivariate GARCH Models by Problem Dimension," KIER Working Papers 815, Kyoto University, Institute of Economic Research.

Chan, L., J. Karceski, and J. Lakonishok (1999): "On Portfolio Optimization: Forecasting Covarainces and Choosing the Risk Model," *The Review of Financial Studies*, 12, 937–974.

Chen, S. (1999): "Beta Kernel Estimators for Density Functions," *Computational Statistics & Data Analysis*, 31, 131–145.

——— (2000): "Probability Density Function Estimation Using Gamma Kernels," *Annals of the Institute of Statistical Mathematics*, 52, 471–480.

Cheng, M., J. Fan, and J. Marron (1997): "On Automatic Boundary Corrections," *The Annals of Statistics*, 25, 1691–1708.

Chiriac, R. and V. Voev (2011): "Modelling and Forecasting Multivariate Realized Volatility," *Journal of Applied Econometrics*, 26, 922–947.

Cipollini, F., R. F. Engle, and G. M. Gallo (2006): "Vector Multiplicative Error Models: Representation and Inference," NYU Working Paper No. Fin-07-048.

Connor, G. (1995): "The Three Types of Factor Models: A Comparison of Their Explanatory Power," *Financial Analysts Journal*, 51, 42–46.

Corradi, V., W. Distaso, and N. R. Swanson (2009): "Predictive Density Estimators for Daily Volatility Based on the Use of Realized Measures," *Journal of Econometrics*, 150, 119–138.

Corsi, F. (2009): "A Simple Approximate Long-Memory Model of Realized Volatility," *Journal of Financial Econometrics*, 7, 174–196.

Dacorogna, M., R. Gencay, U. Muller, R. Olsen, and O. Pictet (2001): *An Intoduction to High-Frequency Finance*, San Diego, CA: Academic Press.

De Luca, G. and G. M. Gallo (2004): "Mixture Processes for Financial Intradaily Durations," *Studies in Nonlinear Dynamics & Econometrics*, 8, –.

De Luca, G. and P. Zuccolotto (2006): "Regime-Switching Pareto Distributions for ACD Models," *Computational Statistics & Data Analysis*, 51, 2179–2191.

de Pooter, M., M. Martens, and D. van Dijk (2008): "Predicting the Daily Covariance Matrix of S&P100 Stocks Using Intraday Data - but which Frequency to Use?" *Econometric Reviews*, 27, 199–229.

DeMiguel, V., L. Garlappi, and R. Uppal (2009): "Optimal versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy?" *Review of Financial Studies*, 22, 1915–1953.

Diebold, F. X., T. A. Gunther, and A. S. Tay (1998): "Evaluating Density Forecasts with Application to Financial Risk Management," *International Economic Review*, 39, 863–883.

Doornik, J. and H. Hansen (2008): "An Omnibus Test for Univariate and Multivariate Normality," *Oxford Bulletin of Economics and Statistics*, 70, 927–939.

Drost, F. C. and B. J. M. Werker (2004): "Semiparametric Duration Models," *Journal of Business and Economic Statistics*, 22, 40–50.

Duan, N. J., W. G. Manning, C. N. Morris, and J. P. Newhouse (1983): "A Comparison of Alternative Models for the Demand for Medical Care," *Journal of Business and Economic Statistics*, 1, 115–126.

Easley, D. and M. O'Hara (1992): "Time and the Process of Security Price Adjustment," *Journal of Finance*, 47, 577–605.

Engle, R. F. (2000): "The Econometrics of Ultra-High-Frequency-Data," *Econometrica*, 68, 1–22.

——— (2002a): "Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroscedasticity Models," *Journal of Business & Economic Statistics*, 20, 339–350.

——— (2002b): "New Frontiers for ARCH Models," *Journal of Applied Econometrics*, 17, 425–446.

——— (2009): *Anticipating Correlations: A New Paradigm for Risk Management*, The Econometric Institute Lectures, Princeton, NJ: Princeton University Press.

Engle, R. F. and G. M. Gallo (2006): "A Multiple Indicators Model for Volatility Using Intra-Daily Data," *Journal of Econometrics*, 131, 3–27.

Engle, R. F. and J. Mezrich (1996): "GARCH for Groups," *Risk*, 9, 36–40.

Engle, R. F. and J. R. Russell (1998): "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data," *Econometrica*, 66, 1127–1162.

Engle, R. F., N. Shephard, and K. Sheppard (2008): "Fitting and Testing Vast Dimensional Time-Varying Covariance Models," Tech. rep., Oxford University.

Engle, R. F. and K. Sheppard (2005): "Theoretical Properties of Dynamic Conditional Correlation Multivariate GARCH," Working paper, University of California, San Diego.

Epps, T. (1979): "Comovement in Stock Prices in the Very Short Run," *Journal of the American Statistical Association*, 74, 291–298.

Fama, E. F. and K. R. French (1993): "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33, 3–56.

Fan, J., Y. Fan, and J. Lv (2008): "High Dimensional Covariance Matrix Estimation Using a Factor Model," *Journal of Econometrics*, 147, 186–197.

Fan, J., Y. Li, and K. Yu (2012a): "Vast Volatility Matrix Estimation using High Frequency Data for Portfolio Selection," *Journal of the American Statistical Association*, 107, 412–428.

Fan, J., J. Zhang, and K. Yu (2012b): "Vast Portfolio Selection With Gross-Exposure Constraints," *Journal of the American Statistical Association*, 107, 592–606.

Fan, Y. (1994): "Testing the Goodness of Fit of a Parametric Density Function by Kernel Method," *Econometric Theory*, 10, 316–356.

——— (1998): "Goodness-of-Fit Tests Based on Kernel Density Estimators with Fixed Smoothing Parameters," *Econometric Theory*, 14, 604–621.

Fernandes, M. and J. Grammig (2005): "Nonparametric Specification Tests for Conditional Duration Models," *Journal of Econometrics*, 127, 35–68.

Fernandes, M. and P. Monteiro (2005): "Central Limit Theorem for Asymmmetric Kernel Functionals," *Annals of the Institute of Statistical Mathematics*, 57, 425–442.

Ferraty, F. and P. Vieu (2006): *Nonparametric Functional Data Analysis: Theory and Practice*, New York: Springer.

Fleming, J., C. Kirby, and B. Ostdiek (2001): "The Economic Value of Volatility Timing," *Journal of Finance*, 56, 329–352.

——— (2003): "The Economic Value of Volatility Timing Using "Realized" Volatility," *Journal of Financial Economics*, 67, 473–509.

Gasser, T. and H. Müller (1979): "Kernel Estimation of Regression Functions," in *Lecture Notes in Mathematics 757*, ed. by T. Gasser and M. Rosenblatt, Heidelberg: Springer, 23–68.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2006): "Predicting Volatility: Getting the Most out of Return Data Sampled at Different Frequencies," *Journal of Econometrics*, 131, 59–95.

Hagmann, M. and O. Scaillet (2007): "Local Multiplicative Bias Correction for Asymmetric Kernel Density Estimators," *Journal of Econometrics*, 141, 213–249.

Hansen, P. R. and A. Lunde (2005): "A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)?" *Journal of Applied Econometrics*, 20, 873–889.

——— (2006): "Realized Variance and Market Microstructure Noise," *Journal of Business and Economic Statistics*, 24, 127–161.

Hansen, P. R., A. Lunde, and V. Voev (2010): "Realized Beta GARCH: A Multivariate GARCH Model with Realized Measures of Volatility and CoVolatility," CREATES Research Papers 2010-74, School of Economics and Management, University of Aarhus.

Härdle, W. K., N. Hautsch, and A. Mihoci (2012): "Local Adaptive Multiplicative Error Models for High-Frequency Forecasts," SFB 649 Discussion Papers SFB649DP2012-031, Sonderforschungsbereich 649, Humboldt University, Berlin, Germany.

Hautsch, N. (2003): "Assessing the Risk of Liquidity Suppliers on the Basis of Excess Demand Intensities," *Journal of Financial Econometrics*, 1, 189–215.

——— (2004): *Modelling Irregularly Spaced Financial Data: Theory and Practice of Dynamic Duration Models*, Berlin: Springer.

——— (2008): "Capturing Common Components in High-Frequency Financial Time Series: A Multivariate Stochastic Multiplicative Error Model," *Journal of Economic Dynamics & Control*, 32, 3978–4009.

Hautsch, N., L. M. Kyj, and P. Malec (2011): "The Merit of High-Frequency Data in Portfolio Allocation," SFB 649 Discussion Papers SFB649DP2011-059, Sonderforschungsbereich 649, Humboldt University, Berlin, Germany.

Hautsch, N., L. M. Kyj, and R. C. A. Oomen (2012): "A Blocking and Regularization Approach to High-Dimensional Realized Covariance Estimation," *Journal of Applied Econometrics*, 27, 625–645.

Hautsch, N., P. Malec, and M. Schienle (2013): "Capturing the Zero: A New Class of Zero-Augmented Distributions and Multiplicative Error Processes," *Journal of Financial Econometrics*, forthcoming.

Hjort, N. L. and I. K. Glad (1995): "Nonparametric Density Estimation with a Parametric Start," *The Annals of Statistics*, 23, 882–904.

Jagannathan, R. and T. Ma (2003): "Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps," *Journal of Finance*, 58, 1651–1683.

Johnson, N. L., S. Kotz, and N. Balakrishnan (1994): *Continuous Univariate Distributions, Volumes I and II*, New York: John Wiley and Sons, second ed.

Jones, M. (1993): "Simple Boundary Correction for Kernel Density Estimation," *Statistics and Computing*, 3, 135–146.

Jones, M. C. and P. J. Foster (1996): "A Simple Nonnegative Boundary Correction Method for Kernel Density Estimation," *Statistica Sinica*, 6, 1005–1013.

J.P. Morgan/Reuters (1996): "RiskMetrics - Technical Document," Tech. rep., J.P. Morgan/Reuters.

Kirby, C. and B. Ostdiek (2012): "Optimizing the Performance of Sample Mean-Variance Efficient Portfolios," Working paper, AFA 2013 San Diego Meetings.

Laloux, L., P. Cizeau, J.-P. Bouchaud, and M. Potters (1999): "Noise Dressing of Financial Correlation Matrices," *Physical Review Letters*, 83, 1467–1470.

Laloux, L., P. Cizeau, M. Potters, and J.-P. Bouchaud (2000): "Random Matrix Theory and Financial Correlations," *International Journal of Theoretical Applied Finance*, 3, 391–397.

Lancaster, T. (1997): *The Econometric Analysis of Transition Data*, Cambridge: Cambridge University Press.

Lanne, M. (2006): "A Mixture Multiplicative Error Model for Realized Volatility," *Journal of Financial Econometrics*, 4, 594–616.

Laurent, S., J. V. K. Rombouts, and F. Violante (2012): "On the Forecasting Accuracy of Multivariate GARCH Models," *Journal of Applied Econometrics*, 27, 934–955.

Ledoit, O. and M. Wolf (2003): "Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection," *Journal of Empirical Finance*, 10, 603–621.

——— (2004): "Honey, I Shrunk the Sample Covariance Matrix," *The Journal of Portfolio Management*, 30, 110–119.

——— (2012): "Nonlinear Shrinkage Estimation of Large-Dimensional Covariance Matrices," *The Annals of Statistics*, 40, 1024–1060.

Liesenfeld, R., I. Nolte, and W. Pohlmeier (2006): "Modelling Financial Transaction Price Movements: A Dynamic Integer Count Model," *Empirical Economics*, 30, 795–825.

Liu, Q. (2009): "On Portfolio Optimization: How and When Do We Benefit from High-Frequency Data?" *Journal of Applied Econometrics*, 24, 560–582.

Lo, A. W. and A. C. MacKinlay (1990): "An Econometric Analysis of Nonsynchronous Trading," *Journal of Econometrics*, 45, 181–211.

Malec, P. and M. Schienle (2012): "Nonparametric Kernel Density Estimation Near the Boundary," SFB 649 Discussion Papers SFB649DP2012-047, Sonderforschungsbereich 649, Humboldt University, Berlin, Germany.

Manganelli, S. (2005): "Duration, Volume and Volatility Impact of Trades," *Journal of Financial Markets*, 8, 377–399.

Markowitz, H. (1952): "Portfolio Selection," *The Journal of Finance*, 7, 77–91.

Marron, J. S. and D. Ruppert (1994): "Transformations to Reduce Boundary Bias in Kernel Density Estimation," *Journal of the Royal Statistical Society. Series B*, 56, 653–671.

Merton, R. C. (1973): "An Intertemporal Capital Asset Pricing Model," *Econometrica*, 41, 867–888.

——— (1980): "On Estimating the Expected Return on the Market: An Exploratory Investigation," *Journal of Financial Economics*, 8, 323–361.

Michaud, R. O. (1989): "The Markowitz Optimization Enigma: Is Optimized Optimal?" *Financial Analysts Journal*, 45, 31–42.

Michels, P. (1992): "Asymmetric Kernel Functions in Non-Parametric Regression Analysis and Prediction," *Journal of the Royal Statistical Society. Series D (The Statistician)*, 41, 439–454.

Noureldin, D., N. Shephard, and K. Sheppard (2012): "Multivariate High-Frequency-Based Volatility (HEAVY) Models," *Journal of Applied Econometrics*, 27, 907–933.

O'Hara, M., C. Yao, and M. Ye (2011): "What's Not There: The Odd-Lot Bias in TAQ Data," Johnson School Research Paper Series 31-2011.

O'Hara, M. and M. Ye (2011): "Is Market Fragmentation Harming Market Quality?" *Journal of Financial Economics*, 100, 459–474.

Oomen, R. C. A. (2009): "High Dimensional Covariance Forecasting for Short Intra-Day Horizons," Working Paper.

Patton, A. and K. Sheppard (2008): "Evaluating Volatility and Correlation Forecasts," in *Handbook of Financial Time Series*, ed. by T. G. Andersen, R. A. Davis, J.-P. Kreiss, and T. Mikosch, Heidelberg: Springer, 801–828.

Robinson, P. and M. Henry (2003): "Higher-Order Kernel Semiparametric M-Estimation of Long Memory," *Journal of Econometrics*, 114, 1–27.

Russell, J. R. and R. F. Engle (2005): "A Discrete-State Continuous-Time Model of Financial Transactions Prices and Times: The Autoregressive Conditional Multinomial-Autoregressive Conditional Duration Model," *Journal of Business & Economic Statistics*, 23, 166–180.

Rydberg, T. and N. Shephard (2003): "Dynamics of Trade-by-Trade Price Movements: Decomposition and Models," *Journal of Financial Econometrics*, 1, 2–25.

Scaillet, O. (2004): "Density Estimation Using Inverse and Reciprocal Inverse Gaussian Kernels," *Journal of Nonparametric Statistics*, 16, 217–226.

Schuster, E. (1958): "Incorporating Support Constraints into Nonparametric Estimators of Densities," *Communications in Statistics, Part A - Theory and Methods*, 14, 1123–1136.

Sharpe, W. (1963): "A Simplified Model for Portfolio Analysis," *Management Science*, 9, 277–293.

Sheppard, K. (2012): "Forecasting High Dimensional Covariance Matrices," in *Handbook of Volatility Models and Their Applications*, ed. by L. Bauwens, C. Hafner, and S. Laurent, Hoboken, New Jersey: Wiley, 103–125.

Stein, J. (1956): "Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution," in *Proceedings of the Third Berkeley Symposium on Mathematical and Statistical Probability*, ed. by J. Neyman, University of California, Berkeley, 197–206.

Tola, V., F. Lillo, M. Gallegati, and R. Mantegna (2008): "Cluster Analysis for Portfolio Optimization," *Journal of Economic Dynamics and Control*, 32, 235–258.

Voev, V. (2009): "On the Economic Evaluation of Volatility Forecasts," CREATES Research Papers 2009-56, School of Economics and Management, University of Aarhus.

Wang, Y. and J. Zou (2010): "Vast Volatility Matrix Estimation for High-Frequency Financial Data," *Annals of Statistics*, 38, 943–978.

Weglarczyk, S., W. G. Stupczewski, and V. P. Singh (2005): "Three Parameter Discontinuous Distributions for Hydrological Samples with Zero Values," *Hydrological Processes*, 19, 2899–2914.

West, K. D., H. J. Edison, and D. Cho (1993): "A Utility-Based Comparison of Some Models of Exchange Rate Volatility," *Journal of International Economics*, 35, 23–45.

Zhang, M. Y., J. R. Russell, and R. S. Tsay (2001): "A Nonlinear Autoregressive Conditional Duration Model with Applications to Financial Transaction Data," *Journal of Econometrics*, 104, 179–207.

Zhang, S. (2010): "A Note on the Performance of Gamma Kernel Estimators at the Boundary," *Statistics and Probability Letters*, 80, 548–557.

Zumbach, G. (2006): "The RiskMetrics 2006 Methodology," Tech. rep., RiskMetrics Group.

——— (2009a): "The Empirical Properties of Large Covariance Matrices," Tech. rep., RiskMetrics Group.

——— (2009b): "Inference on Multivariate ARCH Processes with Large Sizes," Tech. rep., RiskMetrics Group.

Appendix A

A.1 Power of Distribution Tests for Probability Integral Transforms

We simulate 1000 samples of length 8000 considering two DGPs. They are equivalent to DGPs 1 and 2 from the simulation study in Section 1.2.4 with the exception that the constant probability of nonzero observations is replaced by ACM dynamics as in (1.28), (1.31) and (1.32). The autoregressive and moving average parameters ζ_1 and ρ_1 are chosen in line with the estimates obtained in the empirical application. The constant ϖ is specified such that the initial value of π_t equals 0.5 and about 0.9 for DGP 1 and 2, respectively.

For each DGP, we estimate the following models, all assuming the correct specification of the conditional mean μ_t:[1]

- E-MEM: MEM (1.21), where $\mathcal{PMD}(1)$ is based on $\pi = 1$ and $g_\varepsilon(\varepsilon_t) = f_\varepsilon(\varepsilon_t)$ is the standard exponential density.
- G-ZA-MEM: MEM (1.21), where $\mathcal{PMD}(1)$ is based on a constant $\pi_t = \pi$ and $g_\varepsilon(\varepsilon_t)$ is the gamma density with shape parameter m and scale parameter $\lambda = (\pi\, m)^{-1}$.
- ZA-MEM: MEM (1.21), where $\mathcal{PMD}(1)$ is the ZAF density (1.26) with constant $\pi_t = \pi$ and scale parameter $\lambda = (\pi\, \xi)^{-1}$.
- G-LOG-DZA-MEM: MEM (1.21), where $\mathcal{PMD}(1)$ is based on the autologistic model (1.28) and (1.29) with $l = 0$ and $d = 1$ for π_t, while $g_\varepsilon(\varepsilon_t|\mathcal{H}_{t-1})$ is the gamma density with scale $\lambda_t = (\pi_t\, m)^{-1}$.
- LOG-DZA-MEM: MEM (1.21), where $\mathcal{PMD}(1)$ is the ZAF density (1.26) with the autologistic model (1.28) and (1.29), where $l = 0$ and $d = 1$, for π_t, while $\lambda_t = (\pi_t\, \xi)^{-1}$.

[1] Results for two additional DGPs and more estimated models can be found in the web appendix available at `http://amor.cms.hu-berlin.de/~malecpet/ZAMEM_appendix.pdf`.

- G-ACM-DZA-MEM: same as G-LOG-DZA-MEM, but the ACM model in (1.28), (1.31) and (1.32) with $v = w = 1$ is assumed for π_t.

The results of the power study are reported in Table A.1.

Table A.1: Power of Distribution Tests for (Transformed) PITs
Rejection rates of χ^2 and Kolmogorov-Smirnov test for uniformity of PITs, as well as of Bowman-Shenton and Doornik-Hansen test for normality of transformed PITs. Both DGPs assume a DZA-Log-MEM based on a ZAF distribution with $a = 0.6$, $m = 100$, $\eta = 3.3$ and conditional mean parameters $\omega = 0.05$, $\alpha_1 = 0.05$, $\beta_1 = 0.9$, $\alpha_1^0 = -0.005$. π_t follows ACM dynamics with $\rho_1 = 0.15$, $\zeta_1 = 0.99$ and $\varpi = 0.022$ (DGP 1) or $\varpi = 0$ (DGP 2). For every replication, six models are estimated and (randomized) PITs are computed. In the out-of-sample setting, models are estimated using the first two thirds of the sample and PITs are computed based on the remaining third of the dataset. The study uses 1000 replications and a sample size of 8000.

	In-Sample				**Out-of-Sample**			
	DGP 1		**DGP 2**		**DGP 1**		**DGP 2**	
Est. Model\α	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01
	Pearson-χ^2							
E-MEM	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
G-ZA-MEM	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
ZA-MEM	0.025	0.003	0.015	0.003	0.669	0.571	0.721	0.641
G-LOG-DZA-MEM	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
LOG-DZA-MEM	0.047	0.008	0.229	0.088	0.608	0.500	0.711	0.603
G-ACM-DZA-MEM	1.000	1.000	1.000	1.000	1.000	1.000	0.995	0.988
	Kolmogorov-Smirnov							
E-MEM	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
G-ZA-MEM	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
ZA-MEM	0.000	0.000	0.002	0.001	0.706	0.609	0.811	0.764
G-LOG-DZA-MEM	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
LOG-DZA-MEM	0.000	0.000	0.005	0.000	0.651	0.535	0.792	0.713
G-ACM-DZA-MEM	1.000	1.000	1.000	1.000	1.000	1.000	0.997	0.984
	Bowman-Shenton							
E-MEM	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
G-ZA-MEM	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
ZA-MEM	0.015	0.002	0.015	0.001	0.325	0.194	0.499	0.364
G-LOG-DZA-MEM	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
LOG-DZA-MEM	0.144	0.030	0.114	0.018	0.307	0.154	0.439	0.302
G-ACM-DZA-MEM	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Doornik-Hansen							
E-MEM	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
G-ZA-MEM	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
ZA-MEM	0.017	0.002	0.016	0.001	0.333	0.199	0.513	0.365
G-LOG-DZA-MEM	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
LOG-DZA-MEM	0.159	0.044	0.121	0.025	0.316	0.164	0.439	0.312
G-ACM-DZA-MEM	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Appendix B

B.1 MEM Specifications

For trading volumes, we specify the conditional mean $\mu_t^{(v)}$ in (2.34) using the logarithmic MEM proposed by Bauwens and Giot (2000). The latter does not require parameter constraints to ensure the positivity of $\mu_t^{(v)}$ and implies

$$\ln \mu_t^{(v)} = \omega + \sum_{i=1}^{p} \alpha_i \ln x_{t-i}^{(v)} + \sum_{i=1}^{q} \beta_i \ln \mu_{t-i}^{(v)}, \tag{B.1.1}$$

where the lag structure is chosen according to the Schwarz information criterion (SIC).

In case of volatilities, we consider (B.1.1) with $p = 1$, but augmented by the lags of (logarithmic) weekly and monthly realized kernel estimates, which are defined as the averages

$$x_{t,w}^{(rk)} := \frac{1}{5} \sum_{j=0}^{4} x_{t-j}^{(rk)} \quad \text{and} \quad x_{t,m}^{(rk)} := \frac{1}{20} \sum_{j=0}^{19} x_{t-j}^{(rk)}. \tag{B.1.2}$$

This extension is motivated by the widely-used heterogeneous autoregressive (HAR) model for realized volatilities proposed by Corsi (2009) and yields

$$\ln \mu_t^{(rk)} = \omega + \alpha^d \ln x_{t-1}^{(rk)} + \alpha^w \ln x_{t-1,w}^{(rk)} + \alpha^m \ln x_{t-1,m}^{(rk)} + \sum_{i=1}^{q} \beta_i \ln \mu_{t-i}^{(rk)}, \tag{B.1.3}$$

where q is determined using the SIC.

Appendix C

C.1 Analytical Solution for the Performance Fee

Consider the GMV framework (3.1) and the preference structure (3.4). In addition, let

$$\overline{\mu_i^p} := \frac{1}{T-h}\sum_{t=1}^{T-h} \widehat{w}_{t,t+h}^{i'}\,\mu_{t,t+h},\ \overline{\mu_i^{2,p}} := \frac{1}{T-h}\sum_{t=1}^{T-h}\left(\widehat{w}_{t,t+h}^{i'}\,\mu_{t,t+h}\right)^2,\ i=\mathrm{I},\mathrm{II}, \tag{C.1.1}$$

and $\vartheta := 2\,(1+\gamma)\,/\gamma$. Then, exploiting the fact that

$$\mathrm{E}\left[\left(r_{t,t+h}^{p,i}\right)^2\middle|\mathcal{F}_t\right] = \widehat{w}_{t,t+h}^{i'}\,\Sigma_{t,t+h}\,\widehat{w}_{t,t+h}^{i} + \left(\widehat{w}_{t,t+h}^{i'}\,\mu_{t,t+h}\right)^2,\ i=\mathrm{I},\mathrm{II}, \tag{C.1.2}$$

and using basic algebra, condition (3.5) can be rearranged to

$$\Delta_\gamma^2 + \Delta_\gamma\left[\vartheta - 2\left(1+\overline{\mu_{\mathrm{II}}^p}\right)\right] = (\vartheta-2)\left(\overline{\mu_{\mathrm{II}}^p} - \overline{\mu_{\mathrm{I}}^p}\right) + \overline{\mu_{\mathrm{I}}^{2,p}} - \overline{\mu_{\mathrm{II}}^{2,p}} + \overline{\sigma_{\mathrm{I}}^{2,p}} - \overline{\sigma_{\mathrm{II}}^{2,p}}, \tag{C.1.3}$$

where $\overline{\sigma_i^{2,p}}$, $i=\mathrm{I},\mathrm{II}$, is defined as in (3.6). If we assume that $\mu_{t,t+h} = (h/252)\,\mu^{id}\,\iota$, $t=1,\ldots,T-h$, (C.1.3) becomes

$$\Delta_\gamma^2 + \Delta_\gamma\left[\vartheta - 2\left(1+\frac{h\,\mu^{id}}{252}\right)\right] = \overline{\sigma_{\mathrm{I}}^{2,p}} - \overline{\sigma_{\mathrm{II}}^{2,p}}, \tag{C.1.4}$$

yielding the solution

$$\Delta_\gamma = \frac{h\,\mu^{id}}{252} - \frac{1}{\gamma} + \sqrt{\left(\frac{h\,\mu^{id}}{252}-\frac{1}{\gamma}\right)^2 + \overline{\sigma_{\mathrm{I}}^{2,p}} - \overline{\sigma_{\mathrm{II}}^{2,p}}}, \tag{C.1.5}$$

which, under the assumption that $(h/252)\,\mu^{id} \leq 1/\gamma$, is strictly positive only if $\overline{\sigma_{\mathrm{I}}^{2,p}} > \overline{\sigma_{\mathrm{II}}^{2,p}}$.

C.2 Eigenvalue Cleaning

Eigenvalue cleaning is a regularization technique proposed by Laloux et al. (1999) and further developed by Tola et al. (2008) that draws upon random matrix theory to determine the distribution of the eigenvalues of a correlation matrix estimate R depending on the ratio of n observations and m dimensions, $q := n/m$. The idea is to compare empirical correlation eigenvalues with those implied by the null hypothesis of independent Gaussian asset returns, which allows for an identification of those eigenvalues that deviate from the "noisy" ones and hence constitute "signals".

Denote by $\Lambda := \operatorname{diag}(\lambda_1, \ldots, \lambda_m)$ the diagonal matrix of eigenvalues of R ordered from largest to smallest and by Q the matrix of corresponding eigenvectors, yielding the spectral decomposition $R = Q\,\Lambda\,Q'$. For $n \to \infty$, under the null hypothesis R is given by the identity matrix, implying that all eigenvalues are equal to one. However, if $m, n \to \infty$ with $q \geq 1$ fixed, the eigenvalues of R follow a Marchenko–Pastur distribution with maximum eigenvalue $\lambda_{\max} := \big(1 + 1/q + 2\sqrt{1/q}\big)$. Hautsch et al. (2012) argue that, for practical purposes, the above threshold should be tightened to

$$\lambda^*_{\max} := (1 - \lambda_1/m)\,\big(1 + 1/q + 2\sqrt{1/q}\big). \tag{C.2.6}$$

This adjustment allows for a better identification of smaller signals, as it accounts for the fact that the largest empirical eigenvalue λ_1 often is associated with a dominating "market factor". Then, eigenvalue cleaning requires that all eigenvalues below $\lambda^*_{\max}$ are transformed according to

$$\tilde{\lambda}_i := \begin{cases} \lambda_i & \text{if} \quad \lambda_i \geq \lambda^*_{\max}, \\ \delta & \text{otherwise,} \end{cases} \tag{C.2.7}$$

where δ is the average of the positive parts of all "noisy" eigenvalues, i.e.,

$$\delta := \frac{\sum_{(\lambda_i < \lambda^*_{\max})} \lambda_i^+}{\big(\#\text{ of } \lambda_i < \lambda^*_{\max}\big)}. \tag{C.2.8}$$

Finally, the cleaned correlation matrix estimate is obtained as $\tilde{R} = Q\,\tilde{\Lambda}\,Q'$, where $\tilde{\Lambda} := \operatorname{diag}\big(\tilde{\lambda}_i\big)$, $i = 1, \ldots, m$. We apply the procedure to (smoothed) correlation matrix estimates based on the blocked realized kernel, $R^{BRK}_{t,S}$, by setting the number of observations n equal to the minimum number of refresh times in any block averaged over the smoothing window. For the regularization of the rolling window sample covariance of daily returns, C_t, we apply eigenvalue cleaning to the corresponding sample correlation matrix R^c_t with n equal to the window length L.

C.3 Selection of the Number of Factors

To select the number of factors for the regularization approach discussed in Section 3.3.1, we employ the criteria by Bai and Ng (2002) developed for linear factor models with m assets and n observations. In the context of smoothed BRK estimates, we consider a factor model defined in refresh time. Let $r^{(i)}_{t,S,l}$, $i = 1, \ldots, m$, denote the l-th refresh time return from days $t - S + 1$ to t. The resulting factor structure reads

$$r^{(i)}_{t,S,l} = \psi'_{t,S,i} F_{t,S,l} + \varepsilon^{(i)}_{t,S,l}, \; i = 1, \ldots, m, \; l = 1, \ldots, n_{t,S}, \tag{C.3.9}$$

where $F_{t,S,l}$ is the $(k_{t,S} \times 1)$ vector of common factors, $\psi_{t,S,i}$ denotes the corresponding vector of factor loadings and $\varepsilon^{(i)}_{t,S,l}$ is the idiosyncratic component of $r^{(i)}_{t,S,l}$, $i = 1, \ldots, m$. Following Bai and Ng (2002), we determine $k_{t,S}$ by employing the minima of the criteria

$$\begin{aligned} C^{m,1}_{t,S}(k_{t,S}) &= \hat{\sigma}^2_{t,S}(k_{t,S}) + k_{t,S}\, \hat{\sigma}^2_{t,S}(k_{\max}) \left(\frac{m + n_{t,S}}{m\, n_{t,S}} \right) \ln\left(\frac{m\, n_{t,S}}{m + n_{t,S}} \right), \\ C^{m,2}_{t,S}(k_{t,S}) &= \hat{\sigma}^2_{t,S}(k_{t,S}) + k_{t,S}\, \hat{\sigma}^2_{t,S}(k_{\max}) \left(\frac{m + n_{t,S}}{m\, n_{t,S}} \right) \ln\Big[\min\left(\sqrt{m}, \sqrt{n_{t,S}} \right)^2 \Big], \end{aligned} \tag{C.3.10}$$

where $k_{\max}$ is the exogenously fixed maximum number of factors, while

$$\hat{\sigma}^2_{t,S}(k_{t,S}) := \frac{1}{m} \sum_{i=1}^{m} \hat{\sigma}^{2,(i)}_{t,S}(k_{t,S}), \tag{C.3.11}$$

with $\hat{\sigma}^{2,(i)}_{t,S}(k_{t,S})$ being an estimate of the residual variance $\mathrm{V}\big[\varepsilon^{(i)}_{t,S,l}\big]$, $i = 1, \ldots, m$.

In practice, we let $n_{t,S}$ be the minimum number of refresh times in any block of the blocked realized kernel averaged over days $t - S + 1$ to t. Further, we set $\hat{\sigma}^{2,(i)}_{t,S}(k_{t,S})$ equal to the i-th diagonal element of $V^{RK}_{t,S} \left(I_m - \mathcal{Q}_{t,S,(k_{t,S})} \right) V^{RK}_{t,S}$, $i = 1, \ldots, m$, where $V^{RK}_{t,S}$ and $\mathcal{Q}_{t,S,(k_{t,S})}$ are defined as in (3.13) and (3.15), respectively. For the factor structure based on the rolling window sample covariance of daily returns in (3.24), the number of observations is equal to the window length L. The factor residual variance is estimated by $\hat{\sigma}^2_t(k_t) := \frac{1}{m} \sum_{i=1}^{m} \hat{\sigma}^{2,(i)}_t(k_t)$, where $\hat{\sigma}^{2,(i)}_t(k_t)$ is the i-th diagonal element of $\big(V^c_t - \mathcal{Q}^c_{t,(k_t)} \big)$, $i = 1, \ldots, m$.

C.4 Cleaning Procedure for S&P 500 Quote Data

The raw dataset described in Section 3.4.1 is cleaned by performing the following steps:

1. Delete entries with negative bid-ask spreads.
2. Delete entries with non-positive bid or ask prices.
3. Delete entries with non-positive bid or ask sizes.
4. Delete entries with bid-ask spread greater than 1% of the current mid-quote.
5. Delete entries for which the mid-quote price is more than 5 times the median mid-quote on the given day.
6. Delete entries for which the mid-quote price deviated by more than 5 mean absolute deviations from a rolling median (excluding the observation under consideration) of 50 observations (25 observations before and 25 after).

A more detailed discussion of data filtering procedures can be found in Barndorff-Nielsen et al. (2008b).

Selbständigkeitserklärung

Ich bezeuge durch meine Unterschrift, dass meine Angaben über die bei der Abfassung meiner Dissertation benutzten Hilfsmittel, über die mir zuteil gewordene Hilfe sowie über frühere Begutachtungen meiner Dissertation in jeder Hinsicht der Wahrheit entsprechen.

Berlin, 12. Februar 2013

Peter Malec